STIC Search Report
Biotech-Chem Library

TO: Minh-Tam Davis
Location: REM/3A24/3C18
Art Unit: 1642
Monday, May 17, 2004
Case Serial Number: 10/048046

From: Paul Schulwitz
Location: Biotech-Chem Library
REM-1A65
Phone: (571)272-2527

paul.schulwitz@uspto.gov

Search Notes

Examiner Davis,

See attached results.

07/99

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Paul Schulwitz
Technical Information Specialist
STIC Biotech/Chem Library
(571)272-2527

From: Davis, Minh-Tam
Sent: Thursday, May 06, 2004 12:57 PM
To: STIC-Biotech/ChemLib
Subject: Search request for 10/048046

Please search in commercial database, PGPUB, issued patent files, and interference:
1) SEQ ID NO:1, 2.
2) Oligomer search for SEQ ID NO:1,
3) Amino acids 31-103, 303-346, 476-641of SEQ ID NO:2.
Thank you.
MINH TAM DAVIS
ART UNIT 1642
ROOM 3A24, MB 3C18
272-0830

Searcher:_____
Phone:_____
Location:_____
Date Picked Up:_____
Date Completed: _5/7_____
Searcher Prep/Review:_____
Clerical:_____
Online time:_____

TYPE OF SEARCH:
NA Sequences:_____
AA Sequences:_____
Structures:_____
Bibliographic:_____
Litigation:_____
Full text:_____
Patent Family:_____
Other:_____

VENDOR/COST (where applic.)
STN:_____
DIALOG:_____
Questel/Orbit:_____
DRLink:_____
Lexis/Nexis:_____
Sequence Sys.:_____
WWW/Internet:_____
Other (specify):_____

```
          APPLICANT: OLEK, Alexander
          APPLICANT: PIEPENBROCK, Christian
          APPLICANT: BERLIN, Kurt
          TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
          TITLE OF INVENTION: by Assessing DNA Methylation
          FILE REFERENCE: 5013.1012
          CURRENT APPLICATION NUMBER: US/10/204,708
          CURRENT FILING DATE: 2003-05-06
          PRIOR APPLICATION NUMBER: PCT/EP01/03971
          PRIOR FILING DATE: 2001-04-06
          PRIOR APPLICATION NUMBER: DE 10019058.8
          PRIOR FILING DATE: 2000-04-06
          PRIOR APPLICATION NUMBER: DE 10019173.8
          PRIOR FILING DATE: 2000-04-07
          PRIOR APPLICATION NUMBER: DE 10032529.7
          PRIOR FILING DATE: 2000-06-30
          PRIOR APPLICATION NUMBER: DE 10043826.1
          PRIOR FILING DATE: 2000-09-01
          NUMBER OF SEQ ID NOS: 98
          SEQ ID NO 62
          LENGTH: 6801
          TYPE: DNA
          ORGANISM: Artificial Sequence
          FEATURE:
          OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
        US-10-204-708-62

        Query Match        1.6%;   Score 42.4;  DB 4;  Length 6801;
        Best Local Similarity   58.9%;   Pred. No. 1.2;
        Matches   73;  Conservative   0;  Mismatches   51;  Indels   0;  Gaps   0;

        Qy  2556 ATTTTATCTATTTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGAGATTTAAATGT 2615
                      ||||||||      ||||||| |||||||  |||||| |||| ||||||||| | ||||||
        Db  2997 ATAAAATCCAACTTTTTTTCTTTATTACTTTATTATTACTTTTAATATTATAAAATTTAAAATA 2938

        Qy  2616 CACAAACTGTAGCACAAATAATAATTAATTAATTTATAATTTACAAATTGACAAAAAAAAAAAA 2675
                      |||||||||||||  ||||||||||   ||||||||||| |||||||||||||||||||||||
        Db  2937 ATTTCACCTAAACAAAATTAATAAATAAAATAAAATAAAAAAAAATCTAAAAAAAAAAAAAATA 2878

        Qy  2676 AAAA 2679
                      ||||
        Db  2877 AAAA 2874

        RESULT 14
        US-09-790-988-1
        ; Sequence 1, Application US/09790988
        ; Patent No. 6632935
        ; GENERAL INFORMATION:
        ;   APPLICANT: SHIGENOBU, SHUJI
        ;   APPLICANT: WATANABE, HIDEMI
        ;   APPLICANT: HATTORI, MASAHIRA
        ;   APPLICANT: SAKAKI, YOSHIYUKI
        ;   TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
        ;   FILE REFERENCE: 081356/0159
        ;   CURRENT APPLICATION NUMBER: US/09/790,988
        ;   CURRENT FILING DATE: 2001-02-23
        ;   PRIOR APPLICATION NUMBER: JP2000-107160
        ;   PRIOR FILING DATE: 2000-04-07
        ;   NUMBER OF SEQ ID NOS: 7
        ;   SOFTWARE: PatentIn Ver. 2.1
        ;   SEQ ID NO 1
        ;   LENGTH: 640681
        ;   TYPE: DNA
        ;   ORGANISM: Buchnera sp.
        US-09-790-988-1

        Query Match        1.6%;   Score 42.4;  DB 4;  Length 640681;
        Best Local Similarity   50.5%;   Pred. No. 13;
        Matches  103;  Conservative   0;  Mismatches  101;  Indels   0;  Gaps   0;

        Qy  2464 AAGTTTATCAAAAAACATTGTTTCAGGAGGAGGGAGGAGCATAAGTTTACAGCCTACGAGGACGT 2523
                      |||| |||| |   |||||    |  |   |   ||  ||    |      |||||   |||||
```

```
        Db  625138 AAGGCTATCTATTATTACACAAATTTATATTTAAAAAGTAATAATTTACAATGGCCTCTTCAT 625197
                        ||   ||         ||||||    |||||     |    | ||   | ||||||  |||
        Qy  2524 ACACAATATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTTTTATTTTAATAGGT 2583
                       |    ||||      || ||||||  ||||||   |||||   |||| |||||
        Db  625198 GCAAAAATATTATGCAATTGGAAAAGCACTGCTCTTTTTCTTTTAATTATATTAAAAAA 625257
                          |||    || ||||     ||||| ||| ||||||    |  |||| |||||
        Qy  2584 TTGGTGCTTATCTTCTAATAGATTTAAATGTCACAAACTGTAGCACAAATAATATAATT 2643
                         |||  |||||||||    |||||| |||||||||||||||  ||||||||||||||
        Db  625258 AAATTATTTTCCTAAAAAAAAAAAGAAAACAGTGAGAATTTATTGAAAATAATCAATAAAG 625317
                         |||    |||||        |||||    ||     |||| |  |   |||||||||
        Qy  2644 TATAATTTACAAATTGACAAAAAAA 2667
                        |||   |||| ||||
        Db  625318 ACAAAATAAAAAATAATAGAATAA 625341

        RESULT 15
        US-09-293-322C-8
        ; Sequence 8, Application US/09293322C
        ; Patent No. 6231110
        ; GENERAL INFORMATION:
        ;   APPLICANT: Pallas, David C
        ;   APPLICANT: Du, Xianxing
        ;   TITLE OF INVENTION: Coding Sequence for Protein Phosphatase Methylesterase,
        ;   Patent No. 6232110
        ;   TITLE OF INVENTION:  Recombinant DNA Molecules and Methods
        ;   FILE REFERENCE: 105-97
        ;   CURRENT APPLICATION NUMBER: US/09/293,322C
        ;   CURRENT FILING DATE: 1999-04-16
        ;   PRIOR APPLICATION NUMBER:  US 60/082,202
        ;   PRIOR FILING DATE: 1998-04-17
        ;   NUMBER OF SEQ ID NOS: 17
        ;   SOFTWARE: PatentIn Ver. 2.0
        ;   SEQ ID NO 8
        ;   LENGTH: 2409
        ;   TYPE: DNA
        ;   ORGANISM: Mus musculus
        ;   FEATURE:
        ;   NAME/KEY: misc_feature
        ;   LOCATION: (1)..(2409)
        ;   OTHER INFORMATION: N is A, T, G or C.
        US-09-293-322C-8

        Query Match        1.6%;   Score 42.2;  DB 3;  Length 2409;
        Best Local Similarity   61.3%;   Pred. No. 0.75;   Indels   0;  Gaps   0;
        Matches   68;  Conservative   0;  Mismatches   43;  Indels   0;

        Qy  2569 TTTATTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTAAATGTCACAAACTGTAGC 2628
                        |||||   ||| |||| |   |||||  | ||||||    ||  ||||||||||
        Db  2294 TTTGTTTCTTGTTATTTATGATCTTGTTTAAGAAATAATATCTTCCAACCTTTAAA 2353

        Qy  2629 ACAAATAATAATAATTATAATTTACAAATTGACAAAAAAAAAAAAAAAAAAAA 2679
                       |||   |||  ||| |||||||||||||| |||||||||||||||||||||
        Db  2354 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2404

        Search completed: May 14, 2004, 20:10:15
        Job time : 228 secs
```

```
         STRANDEDNESS: single
         TOPOLOGY: linear
         IMMEDIATE SOURCE:
           LIBRARY: FIBRNOT01
           CLONE: 53219
US-08-910-925-2

Query Match       1.6%;  Score 42.6;  DB 3;  Length 2369;
Best Local Similarity  58.1%;  Pred. No. 0.58;
Matches   75;  Conservative  0;  Mismatches  54;  Indels  0;  Gaps  0;

Qy  1639 TTCTGCCACCTGTACTGGGGCTGCACCCGACCGGCTGCTACGGCTGCCTGGCCCCGTTT 1698
         |||||||||||||||||||  |||| |    ||| |   || |  ||| |
Db  1873 TGCTGCCACCTGTACTGGAGCTGCTACTGCGAACTACTGCGACTGCTACTACTTCCACTGC 1814

Qy  1699 TGTGAGCTCAACCTGGTGACAAGTGTCTGGACGGCGGTGCTGAACAACAACAGCTACGAG 1758
         |    ||    |   |  |  ||||  |  ||| |||
Db  1813 TGGAACTACTTCCACTGCTACTGGTTGAACTGCTACTAGAACTGCTACTGCTACTGC 1754

Qy  1759 TCAGACATC 1767
         |  || | |
Db  1753 TTCGACTTC 1745

RESULT 10
US-09-252-991A-2236/c
; Sequence 2236, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2236
;   LENGTH: 966
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2236

Query Match       1.6%;  Score 42.4;  DB 4;  Length 966;
Best Local Similarity  54.5%;  Pred. No. 0.4;
Matches   85;  Conservative  0;  Mismatches  71;  Indels  0;  Gaps  0;

Qy   74 CGGGGATGGTGAATCCCGATGGAGCGGCCCCGAGGAAGGCAAGCAGTCGCCGCCGCCGCAGC 133
        ||   |||| | ||| |    |||||| || ||||  |     |||
Db  578 CGCTGCTGTTCATCGCCATCGATCGCCTCGGGCGGCGAGGAGGGCGAGCCGTCTCCTGGCGGAGG 519

Qy  134 CCTGGGGACGGCTCCTGCGTCTGGGCGCGGAGGAGGGCGAGCCGCACGTCCTCCTGGAGGA 193
        |||  || | |  || |  |   |||
Db  518 CCGCTGCGGGGGTCCGCGAGATGCCGGCGCGCAGGACCTGCGCCGCCGATCCTGGAGGCGCG 459

Qy  194 AGCGGGAGTGGACCATCGGGCGGGAGACGAGGTTGCG 229
        |||  | ||| | ||| ||
Db  458 AGCATGCGGGTGCTGATCGGTCATGGCCGGGGTTGGCG 423

RESULT 11
US-09-252-991A-2607
; Sequence 2607, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
```

```
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2607
;   LENGTH: 1734
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2607

Query Match       1.6%;  Score 42.4;  DB 4;  Length 1734;
Best Local Similarity  54.5%;  Pred. No. 0.55;
Matches   85;  Conservative  0;  Mismatches  71;  Indels  0;  Gaps  0;

Qy   74 CGGGGATGGTGAATCCCGATGGAGCGGCCCCGAGGAAGGCAAGCAGTCGCCGCCGCCGCAGC 133
        ||   |||| | ||| |    |||||| || ||||  |     |||
Db 1212 CGCTGCTGTTCATCGCCATCGATCGCCTCGCAGCAACGACAAGGAGCAGCCGCCTGGCGCGGAGG 1271

Qy  134 CCTGGGGACGGCTCCTGCGTCTGGGCGCGGAGGAGGGCGAGCCGCACGTCCTCCTGGAGGA 193
        |||  || | |  || |  |   |||
Db 1272 CCGCTGGCGGGGTCCGCGAGATGCCGGCGCGCAGGACCTGCGCCGCCGATCCTGGAGGCGCG 1331

Qy  194 AGCGGGAGTGGACCATCGGGCGGGAGACGAGGTTGCG 229
        |||  | ||| | ||| ||
Db 1332 AGCATGCGGGTGCTGATCGGTCATGGCCGGGGTTGGCG 1367

RESULT 12
US-09-252-991A-2445
; Sequence 2445, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2445
;   LENGTH: 2016
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2445

Query Match       1.6%;  Score 42.4;  DB 4;  Length 2016;
Best Local Similarity  54.5%;  Pred. No. 0.6;
Matches   85;  Conservative  0;  Mismatches  71;  Indels  0;  Gaps  0;

Qy   74 CGGGGATGGTGAATCCCGATGGAGCGGCCCCGAGGAAGGCAAGCAGTCGCCGCCGCCGCAGC 133
        ||   |||| | ||| |    |||||| || ||||  |     |||
Db 1454 CGCTGCTGTTCATCGCCATCATCGATCGCCTCGCAGCAACGACAAGGAGCAGCCGCCTGGCGCGGAGG 1513

Qy  134 CCTGGGGACGGCTCCTGCGTCTGGGCGCGGAGGAGGGCGAGCCGCACGTCCTCCTGGAGGA 193
        |||  || | |  || |  |   |||
Db 1514 CCGCTGGCGGGGGTCCGCGAGATGCCGGCGCGCAGGACCTGCGCCGCCGATCCTGGAGGCGCG 1573

Qy  194 AGCGGGAGTGGACCATCGGGCGGGAGACGAGGTTGCG 229
        |||  | ||| | ||| ||
Db 1574 AGCATGCGGGTGCTGATCGGTCATGGCCGGGGTTGGCG 1609

RESULT 13
US-10-204-708-62/c
; Sequence 62, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
```

```
Qy   2533  CCTGCTGCTGCGGAAAACCAGCAGCATTTTATCTATTTTTTATTTTAATAGGTTTGGTGCTT  2592
                ||||   ||||          |||||     |     |  ||||   ||||   ||||
Db    826  TCTTGTTGTGAGCAAAGGATTGTTCTGTTTTAAATTTAATTGACTGTGATTTGGTTGGGT   885

Qy   2593  ATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATAATAATTATAATTTA  2652
                |||    |||           |||     ||      |
Db    886  ATTGGCTATTTTAATTTTACTAAAAAAAAGTGTTCAGTTTTAAAAAAAAAAAAAAAAAAA   945

Qy   2653  CAAATTGACAAAAAAAAAAAAAAAAAAAA  2679
                |||||||||||||||||||||
Db    946  AAAAAAAAAAAAAAAAAAAAAAAAAAA   972
```

RESULT 7
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;   APPLICANT: FLEISCHMAN, Robert D.
;   APPLICANT: WHITE, Owen R.
;   APPLICANT: FRASER, Claire M.
;   APPLICANT: VENTER, John C.
;   TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;   TITLE OF INVENTION: TUBERCULOSIS
;   FILE REFERENCE: 24366-20007.00
;   CURRENT APPLICATION NUMBER: US/09/103,840A
;   CURRENT FILING DATE: 1998-06-24
;   NUMBER OF SEQ ID NOS: 2
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```
Query Match          1.7%;  Score  45.2;  DB 3;  Length 4403765;
Best Local Similarity  52.1%;  Pred. No. 6.9;
Matches  101;  Conservative  0;  Mismatches  93;  Indels  0;  Gaps  0;

Qy     32  GACAGCGGCGGCGGCGGCCAGCCCGGTTCCGGTTCGGCGGCGGCGGGGATGTGAATCCCGA    91
Db 3793359 GGCGGCGGCCGGCGGTGGCGGCGGCGGGTACCTGGTCGGCGGCGGCGGGGACCGGC  3793418

Qy     92  TGGAGCGGCCCCAGGAAGGCAAGCAGCAGTGCGCCCGCCAGCCCTGGGACGGCTCCTGC   151
Db 3793419 GGGGGCCGGCGGAAGAATTCCAGCGGTGGCGCCCACCTCACCGGGGGCACCGGAGGGACC  3793478

Qy    152  GTCTGGGCGGCGAGGAGGAGGGCGAGCGCGCACGTCTCCTCCTGAGGAAGCGGGAGTGGACCATCG  211
Db 3793479 GRCGGGGCCGGCGCGGGCGGGCGCCGGGTGGCTCTACGGCAGCGGCGCCGGCCGGTGCCCGGC  3793538

Qy    212  GGCGGAGACGAGGT   225
Db 3793539 GGCGCCCGGCGGGGCT  3793552
```

RESULT 8
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;   APPLICANT: FLEISCHMAN, Robert D.
;   APPLICANT: WHITE, Owen R.
;   APPLICANT: FRASER, Claire M.
;   APPLICANT: VENTER, John C.
;   TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;   TITLE OF INVENTION: TUBERCULOSIS
;   FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

```
Query Match          1.7%;  Score  45.2;  DB 3;  Length 4411529;
Best Local Similarity  52.1%;  Pred. No. 6.9;
Matches  101;  Conservative  0;  Mismatches  93;  Indels  0;  Gaps  0;

Qy     32  GACAGCGGCGGCGGCGGCCAGCCCGGTTCCGGTTCGGCGGCGGCGGGGATGTGAATCCCGA    91
Db 3802537 GGCGGCGGCCGGCGGTGGCGGCGGCGGGTACCTGGTCGGCGGCGGCGGGGACCGGC  3802596

Qy     92  TGGAGCGGCCCCAGGAAGGCAAGCAGCAGTGCGCCCGCCAGCCCTGGGACGGCTCCTGC   151
Db 3802597 GGGGGCCGGCGGAAGAATTCCAGCGGTGGCGCCCACCTCACCGGGGGCACCGGAGGGACC  3802656

Qy    152  GTCTGGGCGGCGAGGAGGAGGGCGAGCGCGCACGTCTCCTCCTGAGGAAGCGGGAGTGGACCATCG  211
Db 3802657 GGCGGGGCCGGCGCGGGCGGGCGCCGGGTGGCTCTACGGCAGCGGCGCCGGCCGGTGCCCGGC  3802716

Qy    212  GGCGGAGACGAGGT   225
Db 3802717 GGCGCCCGGCGGGGCT  3802730
```

RESULT 9
US-08-910-925-2/c
; Sequence 2, Application US/08910925
; Patent No. 6162601
; GENERAL INFORMATION:
;   APPLICANT: Bandman, Olga
;   APPLICANT: Lal, Preeti
;   APPLICANT: Shah, Purvi
;   TITLE OF INVENTION: HUMAN PININ SPLICE VARIANT
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Incyte Pharmaceuticals, Inc.
;   STREET: 3174 Porter Drive
;   CITY: Palo Alto
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94304
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/910,925
;   FILING DATE: Herewith
;   CLASSIFICATION: 514
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Billings, Lucy J.
;   REGISTRATION NUMBER: 36,749
;   REFERENCE/DOCKET NUMBER: PF-0355 US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 650-855-0555
;   TELEFAX: 650-845-4166
;   TELEX:
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2369 base pairs
;   TYPE: nucleic acid

; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/924,747
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 991 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: SOYBEAN
; IMMEDIATE SOURCE:
; CLONE: SSM.PK0067.G5
US-08-924-747-25

Query Match   1.7%; Score 45.4; DB 3; Length 991;
Best Local Similarity   51.2%; Pred. No. 0.068;
Matches 106; Conservative   0; Mismatches 101; Indels   0; Gaps   0;

Qy   2473 AAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTACAGCCTACAGGACGTACACAATAT 2532
               |||||| | || ||| ||||| ||| | ||| | ||
Db   766 AGAACCATTTTAGAGGTAGGATTCATAATAAGTTAGTATGATTTGTTGGGAAACAATTA 825

Qy   2533 CCTGCTGCTCGGGAAAACCACCAGCCATTTTATCCTATTTTTATTTTAATAGGTTTGGTGCTT 2592
               | | || | | |||| ||||| | ||
Db   826 TCTTGTTGTGAGCAAGGATTGTTCTGTTTTAAATTTAATTAATTGACTGTGATTTGGTTGGGT 885

Qy   2593 ATCTTCTAATAAGATTAAATGTCACAAACTGTAGCACAAATAATAATAATTTATAATTTA 2652
               || | || ||| | |||| || || |||
Db   886 ATTGGCTATTTTAATTTAACTAAAAAAAGTGTTCAGTTTTAAAAAAAAAAAAA 945

Qy   2653 CAAATTGACAAAAAAAAAAAAAAA 2679
               || ||||||||||||||||
Db   946 AAAAAAAAAAAAAAAAAAAAAAAA 972

RESULT 5
US-09-247-373B-25
; Sequence 25, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1108-A
; CURRENT APPLICATION NUMBER: US/09/247,373B
; CURRENT FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 08/924,747
; PRIOR FILING DATE: 1997-09-05
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 25
; LENGTH: 991
; TYPE: DNA
; ORGANISM: SOYBEAN
US-09-247-373B-25

Query Match   1.7%; Score 45.4; DB 3; Length 991;
Best Local Similarity   51.2%; Pred. No. 0.068;
Matches 106; Conservative   0; Mismatches 101; Indels   0; Gaps   0;

Qy   2473 AAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTACAGCCTACAGGACGTACACAATAT 2532
               |||||| | || ||| ||||| ||| | ||| | ||
Db   766 AGAACCATTTTAGAGGTAGGATTCATAATAAGTTAGTATGATTTGTTGGGAAACAATTA 825

Qy   2533 CCTGCTGCTCGGGAAAACCACCAGCCATTTTATCCTATTTTTATTTTAATAGGTTTGGTGCTT 2592
               | | || | | |||| ||||| | ||
Db   826 TCTTGTTGTGAGCAAGGATTGTTCTGTTTTAAATTTAATTAATTGACTGTGATTTGGTTGGGT 885

Qy   2593 ATCTTCTAATAAGATTAAATGTCACAAACTGTAGCACAAATAATAATAATTTATAATTTA 2652
               || | || ||| | |||| || || |||
Db   886 ATTGGCTATTTTAATTTAACTAAAAAAAGTGTTCAGTTTTAAAAAAAAAAAAA 945

Qy   2653 CAAATTGACAAAAAAAAAAAAAAA 2679
               || ||||||||||||||||
Db   946 AAAAAAAAAAAAAAAAAAAAAAAA 972

RESULT 6
US-09-296-715-25
; Sequence 25, Application US/09296715
; Patent No. 6171839
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
; TITLE OF INVENTION: ENZYMES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/296,715
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 991 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: SOYBEAN
; IMMEDIATE SOURCE:
; CLONE: SSM.PK0067.G5
US-09-296-715-25

Query Match   1.7%; Score 45.4; DB 3; Length 991;
Best Local Similarity   51.2%; Pred. No. 0.068;
Matches 106; Conservative   0; Mismatches 101; Indels   0; Gaps   0;

Qy   2473 AAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTACAGCCTACAGGACGTACACAATAT 2532
               |||||| | || ||| ||||| ||| | ||| | ||
Db   766 AGAACCATTTTAGAGGTAGGATTCATAATAAGTTAGTATGATTTGTTGGGAAACAATTA 825

```
Best Local Similarity   3.4%;  Pred. No. 0.0055;
Matches   13;  Conservative 219;  Mismatches 155;  Indels    0;  Gaps    0;

Qy    191 GGAAGCGGGAGTGGACCATCGGGCGGACGAGGTTGCGACCTTTCCTTCCCCAGCAATA 250
              |  |  :::::  ::: :::::: ::: ::: : ::::: ::
Db   1438 GTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1379

Qy    251 AACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGTGACGTGACAC 310
              :: :::        ::::::: ::: ::: : ::: :::
Db   1378 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1319

Qy    311 TGGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGA 370
              ::: ::::::::::::::: ::: :::: ::
Db   1318 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1259

Qy    371 CATGGCCCTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAATCAACCGGAAC 430
              :: :: ::::::: :: ::::: :::::
Db   1258 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1199

Qy    431 ACAACGTGGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGGCATGACACAAGAATCCT 490
              ::: ::::::: :::::::::: :: ::: :::
Db   1198 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1139

Qy    491 TTGAAGCTAACAAGGAAAATGTGTTCCATGGACCCAAAGATACCTCAGGTGCAGTGCAG 550
              ::: : :::::::: ::: :: ::: ::
Db   1138 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1079

Qy    551 GGCGAGGGGCCGATCCCCGGGTCCCTC 577
              ::::::: :|| | |||||
Db   1078 RRRRRRRRRRRRATCGGAAGCTCCCTC 1052

RESULT 2
US-09-621-976-16656
; Sequence 16656, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;   APPLICANT: Dumas Milne Edwards, J.B.
;   APPLICANT: Jobert, S.
;   APPLICANT: Giordano, J.Y.
;   TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;   FILE REFERENCE: GENSET.054PR2
;   CURRENT APPLICATION NUMBER: US/09/621,976
;   CURRENT FILING DATE:  2000-07-21
;   NUMBER OF SEQ ID NOS: 19335
;   SOFTWARE: Patent.pm
; SEQ ID NO 16656
;   LENGTH: 430
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-16656

Query Match     1.7%;  Score 46.2; DB 4; Length 430;
Best Local Similarity  17.4%;  Pred. No.0.027;
Matches   57;  Conservative 126;  Mismatches 144;  Indels    0;  Gaps    0;

Qy   1528 GGAAGCCACGCCCTGTGCACCTGCTTCCAGCCCATGCCCGACCCGGACGGGAGGCGGAGCGC 1587
              ::|:| :: :::::::: ::|:|:::: : :
Db     56 SGRAMCCYYYKKKGSGSCSRAMCCCCYYYCYKSCCSSYKGSYTTKRAMMWKRRSCYYTYSRRR 115

Qy   1588 GAGCGAGACCCGGCGTGTCGCCCCTCAGCAGTGTGCGGTTCGCCTGCCAGCCTTTCTGCCAC 1647
              ::::::: :: ::: |||||  :   :|— :::: ||
Db    116 MYYYNRSYWMMRSMWKGSCCCCSGGSCYKKKKKGKKGSCCMRSYWMCCYYYKRARRMMWK 175

Qy   1648 CTGTACTGGGCTGCACCCGGACCGGCTGCTACGGCTTCGGCCCCGGTTTTGTGAGCTC 1707
              ::: ::::: |||:::: ::: :::|:::::  :::|:
Db    176 GGSCMWYTKRMWRRMCCCCMRRRSMRRRMCMWKGSYTYCYKSSSNMCMARRWKRARGKK 235

Qy   1708 AACCTGGGTGACAAGTGTCTGACGCGGCGTGCTGAACAACAACAGCTACGAGTCAGACATC 1767
              ::|:|||:: : : :::::::|:::| ||| :|: |:::
Db    236 RMCCYTKGGGRMRKYCCMRKKGRRACCTGTTCACCTGCCGTCCTGCCAGAAGGCCTTC 295

Qy   1768 CTGAAGAATTACCTGGCAACCAGAGGTTTGACATGGAAAAAACATGTTGACCGAGAGCTC 1827
```

```
Db    296 ACCTACCAGCGCATGYRAMCSSCMWMKRAKKKKYMMAWSMRMSKYMARRRGSMMCYYY 355
              | ||   ||| :::|:|:  :|:|::  :::|::: :::::|::: ||: : : |:::|
Qy   1828 GTGGCTCTCCAGCCGGGGAGTGTTTCTG 1854
              ::  :  :::  :|  :.::
Db    356 KSMMSKWYKSSGGRAGGGCTTCAATG 382

RESULT 3
US-09-807-258-13
; Sequence 13, Application US/09807258
; Patent No. 6670166
; GENERAL INFORMATION:
;   APPLICANT: E. I. du Pont de Nemours and Company
;   TITLE OF INVENTION: Arthropod Protein Disulfide Isomerases
;   FILE REFERENCE: BB-1253 PCT
;   CURRENT APPLICATION NUMBER: US/09/807,258
;   CURRENT FILING DATE:  2001-06-11
;   PRIOR APPLICATION NUMBER: 60/104,376
;   PRIOR FILING DATE: 1998-10-15
;   NUMBER OF SEQ ID NOS: 32
;   SOFTWARE: Microsoft Office 97
; SEQ ID NO 13
;   LENGTH: 1759
;   TYPE: DNA
;   ORGANISM: Scolopendra canidens DS
US-09-807-258-13

Query Match     1.7%;  Score 45.8;  DB 4;  Length 1759;
Best Local Similarity  49.4%;  Pred. No. 0.073;
Matches  119;  Conservative   0;  Mismatches 122;  Indels    0;  Gaps    0;

Qy   2439 AGGAAAAGTTTCAAGGAGGAGAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGGAAGCGGAG 2498
              || |||| ||||||  ||||| ||| |||| || : |:|| |:||||:|||:||
Db   1514 AGAAAAAAGTTGAATTGTAAAGTAGCAATTTAGAATTTAAAATATTTGTTCAGTAAAGC 1573

Qy   2499 CATAAGTTTACAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACCAGCATT 2558
              ||||||| |||::|||| |||:|| : |||||||| ::|| : |||| :|::||| ::|||
Db   1574 ACAATTTTTTATTTTTAAGGGAATAAAAATGTAAAAATCAATTATGATAATTTAAAATTT 1633

Qy   2559 TTATCTATTTTTTATTTTAATAGGTTTGGTGCTTATCTTCTCAATAAGATTTAAAATGTCAC 2618
              |||  |||  |||| ||| |||| ||| :|||:|||  | :||  |||||||:|:|||
Db   1634 AGAGTGCTTTATGTTGGTCGTACGTATTGCTTTATTGTCTTTATTTCTTTGTAAGATAAAAATGTTGAA 1693

Qy   2619 AAACTGTAGCACAAATAAATAATTTATAATTTACAAATTGACAAAAAAAAAAAAAAAAA 2678
              || | |:|| :  |||||| ||||:| :|||||||:||||  || ||||||||||||||
Db   1694 AATTAATTCGAAATAAATAAAATTTTTTAATAAATTAGTTTTTACAAAAGAAAAAAAAAAAAA 1753

Qy   2679 A 2679
              |
Db   1754 A 1754

RESULT 4
US-08-924-747-25
; Sequence 25, Application US/08924747
; Patent No. 6063570
; GENERAL INFORMATION:
;   APPLICANT: MCCONIGLE, BRIAN
;   APPLICANT: O'KEEFE, DANIEL
;   TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
;   TITLE OF INVENTION: ENZYMES
;   NUMBER OF SEQUENCES: 32
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
;   STREET: 1007 MARKET STREET
;   CITY: WILMINGTON
;   STATE: DELAWARE
;   COUNTRY: UNITED STATES OF AMERICA
;   ZIP: 19898
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: DISKETTE, 3.50 INCH
;   COMPUTER: IBM PC COMPATIBLE
```

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 14, 2004, 12:14:23 ; Search time 215 Seconds
                                      (without alignments)
                                      6914.946 Million cell updates/sec

Title: US-10-048-046-1
Perfect score: 2679
Sequence: 1 aagaattcggcacgaggccg.........acaaaaaaaaaaaaaaaa 2679

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277745446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 Summaries

Database :   Issued Patents NA:*
               1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
               2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
               3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
               4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
               5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
               6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	51.4	1.9	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	46.2	1.7	430	4	US-09-621-976-16656	Sequence 16656, A
	3	45.8	1.7	1759	4	US-09-807-258-13	Sequence 13, Appl
	4	45.4	1.7	991	3	US-08-924-747-25	Sequence 25, Appl
	5	45.4	1.7	991	3	US-09-247-373B-25	Sequence 25, Appl
	6	45.4	1.7	991	3	US-09-296-715-25	Sequence 25, Appl
	7	45.2	1.7	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	8	45.2	1.7	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	9	42.6	1.6	2369	3	US-08-910-925-2	Sequence 2, Appli
c	10	42.4	1.6	966	4	US-09-252-991A-2236	Sequence 2236, Ap
	11	42.4	1.6	1734	4	US-09-252-991A-2607	Sequence 2607, Ap
	12	42.4	1.6	2016	4	US-09-252-991A-2445	Sequence 2445, Ap
c	13	42.4	1.6	6801	4	US-10-204-708-62	Sequence 62, Appl
	14	42.4	1.6	640681	4	US-09-790-988-1	Sequence 1, Appli
	15	42.2	1.6	2409	3	US-09-293-322C-8	Sequence 8, Appli
	16	42.2	1.6	2409	4	US-09-839-497A-8	Sequence 8, Appli
	17	42.2	1.6	4255	5	PCT-US96-02331-14	Sequence 14, Appl
	18	42.2	1.6	4835	5	US-08-386-495-9	Sequence 9, Appli
	19	42.2	1.6	4835	5	PCT-US96-02331-9	Sequence 9, Appli
	20	41.4	1.6	2271	4	US-09-205-258-243	Sequence 243, App
	21	41.4	1.5	2226	4	US-09-205-258-183	Sequence 183, App
	22	41.2	1.5	1037	4	US-09-489-847-112	Sequence 112, App
c	23	41.2	1.5	8093	4	US-10-204-708-31	Sequence 31, Appl
	24	41	1.5	194	4	US-09-621-976-801	Sequence 801, App
	25	41	1.5	1661	4	US-09-439-554-23	Sequence 23, Appl
	26	40.8	1.5	1192	4	US-09-843-376-10	Sequence 10, Appl
c	27	40.8	1.5	26000	4		
	28	40.6	1.5	1637	4	US-09-205-258-178	Sequence 178, App
	29	40.4	1.5	370	4	US-09-376-113-1	Sequence 1, Appli
	30	40.4	1.5	505	4	US-09-621-976-15639	Sequence 15639, A
	31	40.4	1.5	611	4	US-09-376-113-4	Sequence 4, Appli
	32	40.4	1.5	631	4	US-09-376-113-6	Sequence 6, Appli
	33	40.4	1.5	1114	4	US-09-152-060-41	Sequence 41, App
	34	40.4	1.5	1248	4	US-09-489-847-101	Sequence 101, App
	35	40.4	1.5	3572	4	US-09-575-574-3	Sequence 3, Appli
	36	40.2	1.5	1249	4	US-09-461-325-128	Sequence 128, App
	37	40.2	1.5	1249	4	US-10-012-542-128	Sequence 128, App
c	38	40.2	1.5	1260	4	US-09-461-325-93	Sequence 93, Appl
c	39	40.2	1.5	1260	4	US-10-012-542-93	Sequence 93, Appl
	40	40	1.5	450	4	US-09-252-991A-5753	Sequence 5753, Ap
c	41	40	1.5	837	4	US-09-252-991A-5702	Sequence 5702, Ap
c	42	40	1.5	1035	4	US-09-252-991A-5715	Sequence 5715, Ap
	43	40	1.5	1245	4	US-09-252-991A-5725	Sequence 5725, Ap
	44	40	1.5	1342	4	US-09-489-847-89	Sequence 89, Appl
	45	39.8	1.5	593	4	US-09-904-615-59	Sequence 59, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Foley & Lardner
;   STREET: 1800 Diagonal Road, Suite 500
;   CITY: Alexandria
;   STATE: VA
;   COUNTRY: USA
;   ZIP: 22313-0299
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/232,463
;   FILING DATE:
;   CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/07/935,313
;   FILING DATE:
;   APPLICATION NUMBER: EP 91 114 300.6
;   FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
;   NAME: BENT, Stephen A.
;   REGISTRATION NUMBER: 29,768
;   REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703)836-9300
;   TELEFAX: (703)683-4109
;   TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 7218 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   CLONE: pTZgpt-F1s
US-08-232-463-14

        Query Match   1.9%; Score 51.4; DB 1; Length 7218;

```
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 518 BP; 142 A; 127 C; 108 G; 136 T; 0 U; 5 Other;

Query Match             18.2%;  Score 488.8;  DB 4;  Length 518;
Best Local Similarity   98.8%;  Pred. No. 1.9e-107;
Matches    490;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2171 TTCACAGCCCCTGAGGGAAGGGACCCAGGGTCTCCGACAGTGCTCTGGGGTGACTCTT 2230
              ||||||| ||     ||||||||||||||||||||||||||||||||||||||||||
Db        517 TTNNCAGCCCCNTNAGGGAAGGGACGGACCCAGGGTCTCCGACAGTGCTCTGGGGTGACTCTT 458

Qy       2231 CTGTGGAGCTTTTTACCCTCTGAGTGAGACCCTCCCAGAGACCCCCGGGGGCCGCAGCCCG 2290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        457 CTGTGGAGCTTTTTACCCTCTGAGTGAGACCCTCCCAGAGACCCCCGGGGGCGCCCAGCCCG 398

Qy       2291 CCCTCCTGGTGACGCGTGGGCAGGGCTCGTGGTGGCATCAGCAGCAGGAAGCCTTT 2350
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        397 CCCTCCTGGTGACGCGTGGGCAGGGCTCGTGGTGGCATCAGCAGCAGGACGAAGCCTTT 338

Qy       2351 CTGTAACATGCGGCGCGTCCCGCCGAGAGGGGCAGTTTGCTCTTTTGTACATTTTCCGAA 2410
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        337 CTGTAACATGCGGCGCGTCCCGCCGAGAGGGGCAGTTTGCTCTTTTGTACATTTTCCGAA 278

Qy       2411 ACTACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAGTTCAAGGGAGAAGGGCAAGTTTA 2470
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 ACTACAGTTAAAGCAAAAGTCTGTTTTCAGGAAAAGTTCAAGGGAGAAGGGCAAGTTTA 218

Qy       2471 TCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTACAGGAGACTACACAAT 2530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        217 TCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTACAGGAGACTACACAAT 158

Qy       2531 ATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTTTTATTTTAATAGGTTTGGTGC 2590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        157 ATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTTTTATTTTAATAGGTTTGGTGC 98

Qy       2591 TTATCTTCTAATAAGATTAAATGTCACAAACTGTAGCACAAATAATATAATTTATAATT 2650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         97 TTATCTTCTAATAAGATTAAATGTCACAAACTGTAGCACAAATAATAATATAATTTATAATT 38

Qy       2651 TACAAATTGACAAAAA 2666
              ||||||||||||||| ||
Db         37 TACAAATTGACTAAAA 22
```

Search completed: May 14, 2004, 12:44:57
Job time : 1498 secs

```
Qy    743  GTGGTCCCTCGTGTGGCAAGTGATGAAGTCTCCAGCTTGCCTCAGCTCTCCCAGACAGAA  802
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    130  GTGGTCCCTCGTGTGGCAAGTGATGAAGTCTCCAGCTTGCCTCAGCTCTCCCAGACAGAA  189

Qy    803  AGACTGCGTCCTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGA   862
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190  AGACTGCGTCCTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGA   249

Qy    863  AGAAAATGAGAGG--------------------------------AGATGGGGACCTGAACGGG   900
                 ||||||||||||                                ||||||||||||
Db    250  AGAAAATGAGAGGAGGAGAACACATCCCTTTCTCCTGTAGATGGGGACCTTGACCTGAACGGG   309

Qy    901  CAGTTGTTGGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAGGACGTCAGCAGCA  960
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310  CAGTTGTTGGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAGGACGTCAGCAGCA  369

Qy    961  GCGGCTGGGAAGCCAGCAGCAGGAGGAGGACCGTGACGCTGACACGTTCGCGCGGCTGCTACTCG  1020
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370  GCGGCTGGGAAGCCAGCAGCAAGATGGAGGAGGACCGTGACGCTGACACGTTCGCGCGGCTGCTACTCG  429

Qy   1021  CTGCACGACTGCGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCGGCTTGCTACTCG  1080
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    430  CTGCACGACTGCGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCGGCTTGCTACTCG   489

Qy   1081  GGCTGGATGGAGCGCTGTCCGTCCCTGTGTCCTCCACCTGGAGCCGAGCCGGATCTGT   1140
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    490  GGCTGGATGGAGCGCTGTCCGTCCCTGTGTCCTCCACCTGGAGCCGAGCCGGATCTGT   549

Qy   1141  AAAAACCACATCCTCAACAAC  1161
                 |||||||||||||||||||||
Db    550  AAAAACCACATCCTCAACAAC   570
```

RESULT 15
AAH11859/c
ID   AAH11859 standard; cDNA; 518 BP.
XX
AC   AAH11859;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:8694.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
XX   EP1074617-A2.
PN
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000;   2000JP-00118776.
PR   02-MAY-2000;   2000JP-00183767.
PR   09-JUN-2000;   2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 3; SEQ ID NO 8694; 2537pp + Sequence Listing; English.

KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW   antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;
KW   antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW   cerebroprotective; anticonvulsant; antidepressant; gene therapy; vaccine;
KW   autoimmune disorder; multiple sclerosis; allergic condition;
KW   insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW   lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW   central nervous system disorder; Alzheimer's disease; stroke;
KW   Parkinson's disease; Huntington's disease; coagulation disorder;
KW   haemophilia; thrombosis; inflammatory disorder; Crohn's disease; tumour;
KW   infection; depression; psoriasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200021991-A1.
XX
PD   20-APR-2000.
XX
PF   15-OCT-1999;   99WO-US024206.
XX
PR   15-OCT-1998;   98US-0104436P.
XX
PA   (GEMY) GENETICS INST INC.
XX
PI   Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI   Merberg D, Treacy M, Bowman MR;
XX
DR   WPI; 2000-317938/27.
XX
PT   Isolated polynucleotides, and encoded proteins, comprising secreted
PT   expressed sequence tags (sESTs), useful for treating various disorders
PT   such as autoimmune, infectious, and central nervous system disorders.
XX
PS   Claim 1; Page 437; 803pp; English.
XX
CC   AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC   sequence tags (sESTs), isolated from human, mouse, chicken and rat tissue
CC   sources. The sESTs can have a range of activities depending on the
CC   tissues they were isolated from. The activities include: chemotactic;
CC   proliferative; immunomodulatory; haematopoietic; chemokinetic; analgesic;
CC   haemostatic; thrombolytic; antiinflammatory; cytostatic; antibacterial;
CC   antifungal; antiviral; antidiabetic; antiasthmatic; vulnerary; antiulcer;
CC   osteopathic; neuroprotective; nootropic; antipsoriatic; antiparkinsonian;
CC   cerebroprotective; anticonvulsant; and antidepressant. The sESTs can be
CC   used for gene therapy and in vaccines. The sESTs are useful as probes for
CC   the identification and isolation of full-length cDNAs and genomic DNA
CC   molecules which correspond to the sESTs. Proteins encoded by the sESTs
CC   are useful in assays for determining biological activity and raising
CC   antibodies. They may be useful for treatment of autoimmune disorders
CC   (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC   (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC   osteoporosis, osteoarthritis, central nervous system disorders
CC   (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC   disorders (haemophilia, thrombosis) inflammatory disorders (Crohn's
CC   diseae), tumours, bacterial, fungal or viral infections, depression and
CC   psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC   in the exemplification of the present invention
XX
SQ   Sequence 575 BP; 124 A; 168 C; 165 G; 118 T; 0 U; 0 Other;

Query Match   18.5%; Score 495.8; DB 3; Length 575;
Best Local Similarity 94.8%; Pred. No. 4e-109;
Matches 532; Conservative   0; Mismatches   7; Indels   22; Gaps   1;

```
Qy    623  CAACATCGACGTCAGACCTCTTCCCACAGCCTCGGCCTCTTCCACGGAGCCTTCCTG  682
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     10  CAAAGAGGCCTACAGACCTCTTCCCACACGCCTCGGCCTCTTCCACGGAGCCTTCCTG   69

Qy    683  CAGGGCGAGAGCGTTCCTCCAGTTGTGGGTCTGGGGGTGGTGGCATCTCCCTAAAGGAA  742
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     70  CAGGGCGAGAGCGTTCCTCCAGTTGTGGGTCTGGGGGTGGTGGCATCTCCCTAAAGGAA  129
```

```
XX
PR   31-JAN-2000;  2000US-0179065P.
PR   04-FEB-2000;  2000US-0180628P.
PR   28-JUN-2000;  2000US-0214886P.
PR   07-JUL-2000;  2000US-0216647P.
PR   07-JUL-2000;  2000US-0216880P.
PR   11-JUL-2000;  2000US-0217487P.
PR   11-JUL-2000;  2000US-0217496P.
PR   14-JUL-2000;  2000US-0218290P.
PR   26-JUL-2000;  2000US-0220963P.
PR   26-JUL-2000;  2000US-0220964P.
PR   14-AUG-2000;  2000US-0224518P.
PR   14-AUG-2000;  2000US-0224519P.
PR   14-AUG-2000;  2000US-0252267P.
PR   14-AUG-2000;  2000US-0225268P.
PR   14-AUG-2000;  2000US-0225270P.
PR   14-AUG-2000;  2000US-0254447P.
PR   14-AUG-2000;  2000US-0225757P.
PR   14-AUG-2000;  2000US-0225758P.
PR   22-AUG-2000;  2000US-0226868P.
PR   30-AUG-2000;  2000US-0228924P.
PR   01-SEP-2000;  2000US-0229287P.
PR   01-SEP-2000;  2000US-0229343P.
PR   01-SEP-2000;  2000US-0229344P.
PR   01-SEP-2000;  2000US-0229345P.
PR   05-SEP-2000;  2000US-0229509P.
PR   05-SEP-2000;  2000US-0229513P.
PR   08-SEP-2000;  2000US-0231411P.
PR   21-SEP-2000;  2000US-0234223P.
PR   21-SEP-2000;  2000US-0234274P.
PR   25-SEP-2000;  2000US-0349997P.
PR   27-SEP-2000;  2000US-0235834P.
PR   29-SEP-2000;  2000US-0236327P.
PR   29-SEP-2000;  2000US-0363367P.
PR   29-SEP-2000;  2000US-0236388P.
PR   29-SEP-2000;  2000US-0236369P.
PR   29-SEP-2000;  2000US-0236370P.
PR   02-OCT-2000;  2000US-0236802P.
PR   02-OCT-2000;  2000US-0237037P.
PR   02-OCT-2000;  2000US-0237038P.
PR   02-OCT-2000;  2000US-0237039P.
PR   02-OCT-2000;  2000US-0237040P.
PR   13-OCT-2000;  2000US-0399935P.
PR   20-OCT-2000;  2000US-0240960P.
PR   20-OCT-2000;  2000US-0241785P.
PR   20-OCT-2000;  2000US-0241805P.
PR   01-NOV-2000;  2000US-0244617P.
PR   17-NOV-2000;  2000US-0249299P.
PR   08-DEC-2000;  2000US-0251855P.
PR   08-DEC-2000;  2000US-0251868P.
PR   08-DEC-2000;  2000US-0251869P.
XX
PA   (ROSE/)  ROSEN C A.
PA   (RUBE/)  RUBEN S M.
PA   (BARA/)  BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI;  2003-147444/14.
DR   P-PSDB;  ABU55385.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 1; SEQ ID NO 483; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC   human novel polynucleotides of the invention
XX
SQ   Sequence 693 BP; 145 A; 214 C; 197 G; 123 T; 0 U; 14 Other;

Query Match             20.0%;  Score 536.6;  DB 7;  Length 693;
Best Local Similarity   92.5%;  Pred. No. 6.5e-119;
Matches   608;  Conservative   7;  Mismatches  13;  Indels  29;  Gaps   4;

Qy       1044 GCCCTGCATGCACACGTTCTGCGCGCTTGCTACTCGGGCTGGATCGGAGCGGAGCCTCGTCCCT 1103
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 GCCCTGCATGCACACGTTCTGCGCGCTTGCTACTCGGGCTGGATCGGAGCGGAGCCTCGTCCCT 73

Qy       1104 GTGTCCTACCTGCCGCTGTCCCGTGGAGCGGATCTGTAAAAACCACATCCTCAACAACCT 1163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         74 GTGTCCTACCTGCCGCTGTCCCGTGGAGCGGATCTGTAAAAACCACATCCTCAACAACCT 133

Qy       1164 CGTGGAAGCATACCTCATCCAGACAAGAGTCGCAGTGAAGAGATGTGCAAAG 1223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 CGTGGAAGCATACCTCATCCAGACAAGAGTCGCAGTGAAGAGATGTGCAAAG 193

Qy       1224 TATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTT 1283
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 TATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTT 253

Qy       1284 TTCTGATGAAGAAGGGAGTTCAGAGGACCTGTCTGGAGCTGTGCAGTTGACAGTGAGTC 1343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 TTCTGATGAAGAAGGGAGTTCAGAGGACCTGTCTGGAGCTGTGCAGTTGACAGTGAGTC 313

Qy       1344 CTCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTCGTCCTGAGTACAGAAGGCAGGC 1403
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 CTCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTCGTCCTGAGTACAGAAGGCAGGC 373

Qy       1404 GGGCAGCCTCCCACTGCCCAGCACCCGAGGGCAGCCGAGCCAGGAGCCCCACAGGCCCTGGG 1463
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        374 GGGCAGCCTCCCACTGCCCAGCACCCGAGGGCAGCCGAGCCAGGAGCCCCACAGGCCCTGGG 433

Qy       1464 GGATGCACCCTCCACGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTGTGCCCCTCT 1523
              |||||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db        434 GGATGCACCCTCCACGTCCGTCAGCCTGACGACAG---TCAGGATTACGTGTGCCCCTCT 489

Qy       1524 GCAAGGAAGCCACGCCCTGCACCTGCTCTGCACCTGCTT-CAGCCAGCAGTGCGGGTCTGCGGGTCTGCGTCCAGCAGTGTGCGGGTCTGCGTCGACCGGARAGYVGA 1583
              |||||||||||||||||||||||||||||||||||||||||||:|:|||||||||:|:||
Db        490 GCAAGGAAGCCACGCCCTGCACCTGCTCTGCACCTGCTT-CAGCCAGCAGTGCGGGTCTGCGGGTCTGCGTCCAGCAGTGTGCGGGTCTGCGTCGACCGGARAGYVGA 548

Qy       1584 GGGCGAGCAGGAGACCCGCGGTGTCGCCCCTCAGCAGTGTGCGGGTCGCCTCTGCGGGTCTG 1643
              ||||||:|||||||||||||||:|||||                      :||:|:|||:|
Db        549 -CGGCGARCAGGACCCGGACCCGCGGTGKCGCGCC------TCAMCWTTSTG 584

Qy       1644 CCACCTGTACTGGGGTGCACCCGGACCGGCTGCACGGGCTGCTACGGGCCCCGTTTG 1700
              ||||||:||||||:|||||||||||||||||||||||:||||||||||:|||||:
Db        585 CCACCTGTACTGGGGTGCACCCGGACCGGCTGCACGGGCTGCTACGGGNTTGCTGGCCCCGTTTGNG 641

RESULT 14
AAA44336
ID   AAA44336 standard; cDNA; 575 BP.
XX
AC   AAA44336;
XX
DT   21-AUG-2000 (first entry)
XX
DE   Human secreted expressed sequence tag SEQ ID NO:911.
XX
KW   Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
```

Best Local Similarity 92.5%; Pred. No. 6.5e-119;
Matches 608; Conservative 7; Mismatches 13; Indels 29; Gaps 4;

```
Qy  1044  GCCCTGATGCACACGTTCTGCGCGGCTGCTACTCGGCTGGATGGAGCGGCTCGTCCCT  1103
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14  GCCCTGCATGCACACGTTCTGCGCGGCTGCTACTCGGGCTGGATGGAGCGGCTCGTCCCT  73

Qy  1104  GTGTCCTACCTGCCGCTGTCCCGTGGAGCGGATCTGTAAAAACCACATCCTCAACAACCT  1163
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    74  GTGTCCTACCTGCCGCTGTCCCGTGGAGCGGATCTGTAAAAACCACATCCTCAACAACCT  133

Qy  1164  CGTGGAAGCATACCTCATCCAGCATCCAGACAAGAGTCGCAGTGAACAAGATGTGCAAAG  1223
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   134  CGTGGAAGCATACCTCATCCAGCATCCAGACAACAAGAGTCGCAGTGAAGAAGATGTGCAAAG  193

Qy  1224  TATGGATGCCAGGAATAAAATCACTCAAGACATGCTCAGACCCCAAAGTCAGGCGCGTCTT  1283
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   194  TATGGATGCCAGGAATAAAATCACTCAAGACATGCTCAGCCCAAAGTCAGGCGGGTCTT  253

Qy  1284  TTCTGATGAAGAGGGAGTTCAGAGACCTGCTGGAGCGTGTCAGACGTTGACAGTGAGTC  1343
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   254  TTCTGATGAAGAGGGAGTTCAGAGACCTGCTGGAGCGTGTCAGACGTTGACAGTGAGTC  313

Qy  1344  CTCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAGGCAGGC  1403
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   314  CTCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAGGCAGGC  373

Qy  1404  GGGCAGCCTCCCACTGCCCAGCACCGAGGGGAGCCAGGAGCCCCACAGGCCCTGG  1463
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   374  GGGCAGCCTCCCACTGCCCAGCACCGAGGGGAGCCAGGAGCCCCACAGGCCCTGGG  433

Qy  1464  GGATGCACCCTCACGTCCGTCGCCTGAGCCAGCAGCAGTCCAGGATTACGTGTGCCCTCT  1523
                |||||||||||||||||||||||||||||||           |||||||||||||||
Db   434  GGATGCACCCTCCACGTCCGTCCAGCCTGACGACAG----TCAGGATTACGTGTGCCCTCT  489

Qy  1524  GCAAGGAAGCAGCCACGCCCGTCGTTCCAGCCATGCCCGACCCGGAGCCGGA  1583
                |||||||||||||||||||||||||||||||||||||||:|||||
Db   490  GCAAGGAAGCCAGCCCGTGTGCCACCTGCGCTT-CAGCCCATGCCCGACCCGGACRAGYGGA  548

Qy  1584  GCGGCAGCAGGACCCGCGTGTGCGCCCCTCAGCAGTGTGCGGTCTGCCTGCAGCCTTTCTG  1643
                ||||:||||||||||||||:||||              :|:|:||
Db   549  -CGGARCAGGACCCGCGTGKCGCCC----------------TCAMCWTTSTG  584

Qy  1644  CCACCTGTACTGGGGCTGCACCCCGACCCGACCCGGCTGCTTACGGCCCCGGTTTG  1700
                |:||||||:|||||||||||||||:||||||:|||
Db   585  CCACCTGTACTGGGGCTGCACCCCGACCCGACCCGCTACGNTTGCTGGCCCCGTTTGNG  641
```

```
RESULT 13
ABX73645
ID   ABX73645  standard; DNA; 693 BP.
XX
AC   ABX73645;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polynucleotide #473.
XX
KW   Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
```

```
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251888P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   P-PSDB; AAU16317.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 1; SEQ ID NO 483; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence encodes a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed
```

Query Match   20.0%;   Score 536.6;   DB 4;   Length 693;

```
Db      598  TCCCTCTGTGGCAAGTGATGATGAAGTCTCCAGCTTTGCCTCAGCTCTCCCAGACACAGAAAGAC  657
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      807  TGCGTCCTTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTGAGCCCGTGAGGAAGAA  866
                  ||||||||||||||||||||||||||||||||||||||||
Db      658  TGCGTCCTTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTGGAGCCCGAGGAAGAGAA  717
                  ||||||||||||||||||||||||||||||||||||||||
Qy      867  AATGAGGAGGATGGGGACCTTGACCTGAACGGGCAGTTGGTCGTCCACAACCGGCGTAG  926
                  ||||||||||||||||||||||||||||||||||||||||
Db      718  AATGAGGAGGATGGGGACCTTGACCTGAACGGGCAGTTGTTCGTCGCACAACCGGCGTAG  777
                  ||||||||||||||||||||||||||||||||||||||||
Qy      927  AAATGCCAAACCGTCCACGAGGACGTCAGGAGCGAGGCTGGGAAGC  973
                  ||||||||||||||||||||||||||||||||||||||||
Db      778  AAATGCCAAACCGTCCACGAGGACGTCAGGAGCAGCACNGCTGGGAAGC  824
```

RESULT 12
AAS26304
ID   AAS26304 standard; cDNA; 693 BP.
XX
AC   AAS26304;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human cDNA encoding a novel secreted protein, Seq ID 483.
XX
KW   Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaenia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0242223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0235327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   02-OCT-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0241826P.
PR   08-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
```

```
CC        of an oligonucleotide comprising a sequence complementary to the
CC        complementary strand of a polynucleotide which comprises a 5'-end
CC        sequence and an oligonucleotide comprising a sequence complementary to a
CC        polynucleotide which comprises a 3'-end sequence, where the
CC        oligonucleotide comprises at least 15 nucleotides and the combination of
CC        the 5'-end sequence/3'-end sequence is selected from those defined in the
CC        specification. The primer sets can be used in antisense therapy and in
CC        gene therapy. The primers are useful for synthesising polynucleotides,
CC        particularly full-length cDNAs. The primers are also useful for the
CC        detection and/or diagnosis of the abnormality of the proteins encoded by
CC        the full-length cDNAs. The primers allow obtaining of the full-length
CC        cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC        AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC        represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC        oligonucleotides, all of which are used in the exemplification of the
CC        present invention
XX
SQ        Sequence 824 BP; 200 A; 203 C; 258 G; 161 T; 0 U; 2 Other;

Query Match             21.8%;  Score 583;  DB 4;  Length 824;
Best Local Similarity   80.5%;  Pred. No. 4.7e-130;
Matches  762;  Mismatches   62;  Conservative   0;  Indels 123;  Gaps   2;

Qy        27 CTCCTGACAGCGGCGCGGCGGGCGGCCGCAGCCGGTTCGGCGGGCGGGGGGGGGATGTGAAT   86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 CTCTTGACAGCGGCGCGGCGGGCGGCCGCAGCCGGTTCGGCGGGCGGGGGGGGGATGTGAAT   60

Qy        87 CCCCATGGAGCGGCCCGAGGAAGGCAAGCAGTCGCGGCCGCCGCCCAGCCCTGGGGACCGCT  146
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CCCGATGGAGCGGCCCGAGGAAGGCAAGCAGTCGCGGCCGCCGCCCAGCCCTGGGGACCGCT  120

Qy       147 CCTGCGTCTGGGCGTCTGGGCGCGGAGGAGGGCGAGCCGACAGTCTCCTCTGAGGAGCGGAGTGGAC  206
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CCTGCGTCTGGGCGTCTGGGCGCGGAGGAGGGCGAGCCGACAGTCTCCTCTGAGGAGCGGAGTGGAC  180

Qy       207 CATCGGGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGA  266
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 CATCGGGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGA  240

Qy       267 TCACTGTAGAATTGTAGTGGATGAAAATCAGTCAGTGACACTGAAGATACCAGCAC  326
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TCACTGTAGAATTGTAGTGGATGAAAATCAGTCAGTGACACTGAAGATACCAGCAC  300

Qy       327 CAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAACAGACATGCCCTTTACAGAC  386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 CAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAACAGACATGCCCTTTACAGAC  360

Qy       387 TGGGGATGTCATCTACTTCGTGTACAGGAAGAATGAACCGGAACACACGTGCCATACCT  446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 TGGGGATGTCATCTACTTCGTGTACAGGAAGAATGAACCGGAACACACGTGCCATACCT  420

Qy       447 CTATGAATCTTTAAGTGAAAAGCAAGGCATGACCAAGCAACACAAGAATCCTTTGAAGTGTGCCTTG  506
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 CTATGAATCTTTAAGTGAAAAGCAAGGCATGACCAAGCAACACAAGAATCCTTTGAAGTGTGCCTTG  480

Qy       507 AAATGTTCCATGGACCAAAGATACCTCAGGTGCAGTGCAGGGCGGAGGGGCCGATCC  566
              |||||||                                ||||
Db       481 CTGTTGCCCAGCGGCTGGTCTAAAACTCCTGGGATCAAGTGA--------------  522

Qy       567 CCGGGTCCCTCCGTCGTCGCCCGCCACTCAGGTGTGCTTTGAGGAACCACAGCCATCAAC  626
                 |||||
Db       523 ----TCCTCCCACCTTGGCTCCCAAAGTATTGTGATTACAG--------------------  560

Qy       627 ATCGACGTCAGACCTCTTCCCACGGAGCCTTCTCCTCGCAGG  686
Db       561 ------------------------------------------  560

Qy       687 GCGAGAGCGGTTCCTCCAGTGTGGGTCTCGGGGGTGGCATCTCCCTAAAGGAAGTGG  746
              |||||
Db       561 --------------GGTCTGGGGGTGGTGGTGGTGGCATCTCCCCCCTAAAGGAAGTGG  597

Qy       747 TCCCTCTGTGGCAAGTGATGAAGTTCCTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAAGAC  806
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 1; SEQ ID NO 3661; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 816 BP; 182 A; 216 C; 253 G; 162 T; 0 U; 3 Other;

Query Match           27.1%;     Score 727;      DB 4;    Length 816;
Best Local Similarity 93.5%;     Pred. No. 9.4e-165;
Matches  793;    Conservative     0;   Mismatches  18;   Indels  37;   Gaps   2;

Qy   17 GCCGCAATGTCTCCTTGACAGCGGCGGCGGCGGCGGCGGCCGGTTCCGGCGGCGGCGGCCG 76
        ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5 GCCGCAATGTCTCTTGACAGCGGCGGCGGCGGCGGCGGCGGCCGGTTCCGGCGGCGGCGGCCG 64

Qy   77 GGATGTGAATCCCGATGGAGCGGCCGCGGAAGCCAAGCAGCAGTCGCCGCCGCCGCAGCCCT 136
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 GGATGTGAATCCCGATGGAGCGGCCGCGGAAGCCAAGCAGCAGTCGCCGCCGCCGCAGCCCT 124

Qy  137 GGGGACGGCTCCTGCGTCTGGGCCGGCGGAGGAGGGGCGGAGGTCCGCACGTCCTCCTGAGGAAGC 196
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  125 GGGGACGGCTCCTGCGTCTGGGCCGGCGGAGGAGGGGCGGAGGTCCGCACGTCCTCCTGAGGAAGC 184

Qy  197 GGGAGTGGACCATCGGGCGGCAGACGAGGTTGCGACCTTTCTTCCTTCCCCAGCAATAAACTGG 256
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  185 GGGAGTGGACCATCGGGCGGCAGACGAGGTTGCGACCTTTCTTCCTTCCCCAGCAATAAACTGG 244

Qy  257 TCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGTGACACTGGAAG 316
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  245 TCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGTGACACTGGAAG 304

Qy  317 ATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTAAGAGCAGACATGCC 376
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  305 ATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTAAGAGCAGACATGCC 364

Qy  377 CTTTACAGACTGGGGATGTCATCATCTGGTTACGTACAGGAACGCGGAACCGGAACACAACG 436
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  365 CTTTACAGACTGGGGATGTCATCATCTGGTTACGTACAGGAACGCGGAACCGGAACACAACG 424

Qy  437 TGGCATACCCTATGAATCTTAAGTGAAAAGCAAGGCATGACAACAAGAATCCTTTGAAG 496
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  425 TGGCATACCCTATGAATCTTAAGTGAAAAGCAAGGCATGACAACAAGAATCCTTT---- 480

Qy  497 CTAACAAGAAAAATGTGTTCCATGGACCAAAGATACCTCAGGTGCAGGGCCGAG 556
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   911 GCCGCACTCAGGTGAAAGCTCACCACGCGCCATGAAATTCAATCATATCTGTGAACAGACAA 970

Qy  2072 GGTTCAAAAACTAAGCATCCAGAGCCCTGAGCAGCTTTCAGCACTGGAGGTGAAGAGAG 2131
         ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   971 GGTTCAAAAACTAAGCATCCAGAGCCCTGAGCAGCTTTCAGCACTGGAGGTGAAGAGAG 1030

Qy  2132 CGTGTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTCACAGCCCCCTGAGGGAAG 2191
         ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  1031 CGTGTTTTAAAATACAGAGACAAACACGTCAAGGTGTTTCACAGCCCCCTGAGGGAAG 1090

Qy  2192 GGACGCAGGGGTCTCCGACAGTGTCTCTGGGGTGACTCTTCTGTGGAGCTTTTTACCCTCT 2251
         ||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db  1091 GGACGCAGGGGTCTCCGACAGTGTCTCTGGGGTGACTCTTCTGTGGAGC--TTTACCCTCT 1148

Qy  2252 GAGTGAGACCCTCCCAGAGCCCCGGGGGCCGCAGCCCCTCCTGGTGAGCGGCTGGGC 2311
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1149 GAGTGAGACCCTCCCAGAGCCCCGGGGGCCGCAGCCCCTCCTGGTGAGCGGCTGGGC 1208

Qy  2312 AGGGCTCGTGGTGGCATCAGCAGCAGAGACGAAGCCTTTCTGTAACATGGGCCGTCCCG 2371
         ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db  1209 AGGGCTCGTGGTGGCATCAGCAGCAGAGACGAAGCCTTTCTGTAACATGCGGCCGTCCCG 1268

Qy  2372 CCGAGAGGGGGCAGTTTGCTCTTTGTACATTTTCCGAAACTACAGTTAAAGCAGAAGTC 2431
         ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  1269 CCGAGAGGGGGCAGTTTGCTCTTTTGTACATTTTCCGAAACTACAGTTAAAGCAGAAGTC 1328

Qy  2432 TGTTTTCAGGAAAAGTTTCAAGGAGGAGGAAGGCAAGTTTATCAAAAACATTGTTTCAGGAG 2491
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1329 TGTTTTCAGGAAAAGTTTCAAGGAGGAGGAAGGCAAGTTTATCAAAAACATTGTTTCAGGAG 1388

Qy  2492 AAGGGAGCATAAGTTTACAGCCTACAGGACGTACACAAATATCCTGCTGCTGGGAAAACCA 2551
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1389 AAGGGAGCATAAGTTTACAGCCTACAGGACGTACACAAATATCCTGCTGCTGGGAAAACCA 1448

Qy  2552 CAGCATTTATCTATTTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAA 2611
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1449 CAGCATTTATCTATTTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAA 1508

Qy  2612 ATGTCACAAACTGTAGCACAAATAATATATATTTTATAATTTACAAATTGACAAAA 2666
         ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db  1509 ATGTCACAAACTGTAGCACAAATAATATATATTTTATAATTTACAAATTGACTAAAA 1563
```

```
RESULT 10
AAH06828
ID   AAH06828 standard; cDNA; 816 BP.
XX
AC   AAH06828;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:3663.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000;   2000JP-00118776.
PR   02-MAY-2000;   2000JP-00183767.
PR   09-JUN-2000;   2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
```

```
RESULT 9
ADA52592
ADA52592
ID   ADA52592 standard; cDNA; 2186 BP.
XX
AC   ADA52592;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human coding sequence, SEQ ID 160.
XX
KW   Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW   Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW   inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS   Homo sapiens.
XX
PN   EP1293569-A2.
XX
PD   19-MAR-2003.
XX
PF   21-MAR-2002; 2002EP-00006586.
XX
PR   14-SEP-2001; 2001JP-00328381.
PR   24-JAN-2002; 2002US-0350435P.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI   Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI   Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR   WPI; 2003-395539/38.
DR   P-PSDB; ADA54231.
XX
PT   New polynucleotides encoding full-length polypeptides, e.g. secretory
PT   and/or membrane proteins, useful for developing medicines for diseases in
PT   which the gene is involved, or as target molecules for gene therapy.
XX
PS   Claim 1; SEQ ID NO 160; 205pp; English.
XX
CC   The present invention relates to novel human secretory or membrane
CC   proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC   ADA54071). The coding sequences are useful in the gene therapy of
CC   diseases caused by abnormalities of the proteins, e.g. cancer,
CC   inflammatory diseases, osteoporosis or neurological disease.
XX
SQ   Sequence 2186 BP; 493 A; 555 C; 539 G; 599 T; 0 U; 0 Other;

Query Match      32.7%;     Score 876.2; DB 7;  Length 2186;
Best Local Similarity   99.4%;   Pred. No. 1.7e-200;
Matches  890;  Conservative   0;  Mismatches   3;  Indels   2;  Gaps   1;
```

```
Db   383  --------------------------------------------------------------  382

Qy   596  AGGTGTGCTTTGAGGAACCACAGCCATCAACATCGACGTCAGACCTCTTCCCACCACAGCCT  655

Db   383  ------------------------------------------------------------  382

Qy   656  CGGCCTCTTCCACGGAGCGCCTTCTCCTGCACGGGCGAGAGCCGTTCCCCAGTTGTGGGTCTG  715

Db   383  ------------------------------------------GGTCTG  388
                                                   |||||||
Qy   716  GGGGTGGTGGCATCTCCCTAAAGGAAGTGGTCCCTCGTGGCAAGTGATGAAGTCTCCA  775
         |||||||||||||||||||||||||||||    ||||||||||||||||||
Db   389  GGGGTGGTGGCATCTCCCTAAAGGAAGTGGTGTCCCCTGTGGCAAGTGATGAAGTCTCCA  448

Qy   776  GCTTTGCCTCAGCTCTCCCAGACACAGAAAGACTGCGTCCTTTTCGTCGTTGGAACCCCAGG  835
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   449  GCTTTGCCTCAGCTCTCCCAGACACAGAAAGACTGCGTCCTTTTCGTCGTTGGAACCCCAGG  508

Qy   836  ATCAGGAGGATTTGGAGCCCGTGAACGAAGAAAATGAGAGAGATGGGGACCTTGACCTGA  895
         ||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db   509  ATCAGGAGGATTTGGAGCCCGTGAAGAAGAAATGAGAGAGATGGGGACCTTGACCTGA  568

Qy   896  ACGGGCAGTTGTTGGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAGGACGTCA  955
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   569  ACGGGCAGTTGTTGGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAGGACGTCA  628

Qy   956  GAGCAGCACGCGGCTGGGAGGCCAGACAAGATGAGGAGGAGCGCTGACATGCATGCATCATCTGCCAGG  1015
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   629  GAGCAGCGGCGCTGGGAGGCCAGACAAGATGAGGAGGAGCGCTGACATGCATGCATCATCTGCCAGG  688

Qy  1016  ACCTGCTGCACCACTGCGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCGGCTTGCT  1075
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   689  ACCTGCTGCACCACTGCGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCGGCTTGCT  748

Qy  1076  ACTCGGGCTGGATGAGACGCTCGTCGTCCCTGTGTCCTACCTGCCGCTGTCCGTGGAGCGGA  1135
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   749  ACTCGGGCTGGATGAGACGCTCGTCGTCCCTGTGTCCTACCTGCCGCTGTCCGTGGAGCGGA  808

Qy  1136  TCTGTAAAAACCACCATCCTCAACAACCTCGTTGGAAGCATATCCTCATCCAGCATCCAGACA  1195
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   809  TCTGTAAAAACCACCATCCTCAACAACCTCGTTGGAAGCATATCCTCATCCAGCATCCAGACA  868

Qy  1196  AGAGTCGCAGTGAAGAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACA  1255
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   869  AGAGTCGCAGTGAAGAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACA  928

Qy  1256  TGCTGGAGCCCAAAGTCAGGCGGTCTTTTCTGATGAAGAAGGAGTTCAGAGGACCTGC  1315
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   929  TGCTGGAGCCCAAAGTCAGGCGGTCTTTTTCTGATGAAGAAGGAGTTCAGAGGACCTGC  988

Qy  1316  TGGAGCTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTGCC  1375
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   989  TGGAGCTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTGCC  1048

Qy  1376  GGCAGTGTCCTGAGTACAGAGAGGCAGGCGGCGGGATGCAGCCTCCCACTGCCCCAGCACCCGAGG  1435
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1049  GGCAGTGTCCTGAGTACAGAGAGGCAGGCGGCGGGATGCAGCCTCCCACTGCCCCAGCACCCGAGG  1108

Qy  1436  GCGAGGCCAGGAGCCCCACCAGGCCCTGGGGATGCACCCCTCCACGGTCCGTCGACGA  1495
         ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  1109  GCGAGGCCAGGAGCCCCACCAGGCCCTGGGGATGCACCCCTTCACGGTCCGTCGACGA  1168

Qy  1496  CAGCAGTCCAGATTACGTGTGCCCTCTGCAAGGAGCCACGCCCTGTGCACCTGCTGCT  1555
         |||||||||||    |||||||||||||||||||||||||||||||||||||||||||
Db  1169  CAGT---CAGATTACGTGTGCCCTCTGCAAGGAGCCACGCCCTGTGCACCTGCTGCT  1224

Qy  1556  TCCAGCCCATGCCACCGGAGCCGGAGGCCGGAGCCGGAGGA-CCCGCGTGTCGCCCCTCAG  1614
         ||||||||||||||||||||||||||||||| |||||  ||| : ::|||||||||||
Db  1225  T-CAGCCCATGCCGACGCCGGARAGTGGAACGCCGAACGGGAGCCCCGGTKGSCCCCTNAA  1283

Qy  1615  CA  1616
         ||
Db  1284  CA  1285
```

```
Qy  1772  AGAATTACCTGGCAACCAGGAGTTGACATGGAAAAACATGTTGACCGAGAGCCTCGTGG  1831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   671  AGAATTACCTGGCAACCAGGAGTTGACATGGAAAAACATGTTGACCGAGAGCCTCGTGG  730

Qy  1832  CTCTCCAGCGGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACCGGAGACACCGTTCTGT  1891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   731  CTCTCCAGCGGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACCGGAGACACCGTTCTGT  790

Qy  1892  GTTACTGCTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTC  1951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   791  GTTACTGCTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTC  850

Qy  1952  CTGCTTCCGAGTTGCCAGTGGCCAGTAACATCCCGTCCTGACTGCTGCTACTGGGGCCGTAACT  2011
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   851  CTGCTTCCGAGTTGCCAGTGGCCAGTAACATCCCGTCCTGACTGCTGCTACTGGGGCCGTAACT  910

Qy  2012  GCCGGCACTCAGGTGAAAGCTGAAAGTCCACCACGCACTGGGCCCATGAATCAAGACAA  2071
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
XX
AC    ABX73184;
XX
DT    18-MAR-2003  (first entry)
XX
DE    Human novel polynucleotide #12.
XX
KW    Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW    muscular disorder; respiratory disease; reproductive disorder;
KW    gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW    hyperproliferative disorder; inflammatory disease; allergic reaction;
KW    blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW    cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW    haemostatic; antiarteriosclerotic.
XX
OS    Homo sapiens.
XX
PN    US2002132753-A1.
XX
PD    19-SEP-2002.
XX
PF    17-JAN-2001; 2001US-00764864.
XX
PR    31-JAN-2000; 2000US-0179065P.
PR    04-FEB-2000; 2000US-0180628P.
PR    28-JUN-2000; 2000US-0214886P.
PR    07-JUL-2000; 2000US-0216647P.
PR    07-JUL-2000; 2000US-0216880P.
PR    11-JUL-2000; 2000US-0217487P.
PR    11-JUL-2000; 2000US-0217496P.
PR    14-JUL-2000; 2000US-0218290P.
PR    26-JUL-2000; 2000US-0209963P.
PR    26-JUL-2000; 2000US-0220064P.
PR    14-AUG-2000; 2000US-0224519P.
PR    14-AUG-2000; 2000US-0225267P.
PR    14-AUG-2000; 2000US-0225268P.
PR    14-AUG-2000; 2000US-0225270P.
PR    14-AUG-2000; 2000US-0225447P.
PR    14-AUG-2000; 2000US-0225757P.
PR    14-AUG-2000; 2000US-0225758P.
PR    22-AUG-2000; 2000US-0226868P.
PR    30-AUG-2000; 2000US-0228924P.
PR    01-SEP-2000; 2000US-0229287P.
PR    01-SEP-2000; 2000US-0229343P.
PR    01-SEP-2000; 2000US-0229344P.
PR    01-SEP-2000; 2000US-0229345P.
PR    05-SEP-2000; 2000US-0229509P.
PR    05-SEP-2000; 2000US-0229513P.
PR    08-SEP-2000; 2000US-0231413P.
PR    21-SEP-2000; 2000US-0234223P.
PR    21-SEP-2000; 2000US-0234274P.
PR    25-SEP-2000; 2000US-0234997P.
PR    27-SEP-2000; 2000US-0235834P.
PR    29-SEP-2000; 2000US-0236327P.
PR    29-SEP-2000; 2000US-0236367P.
PR    29-SEP-2000; 2000US-0236368P.
PR    29-SEP-2000; 2000US-0236369P.
PR    29-SEP-2000; 2000US-0236370P.
PR    02-OCT-2000; 2000US-0236802P.
PR    02-OCT-2000; 2000US-0237037P.
PR    02-OCT-2000; 2000US-0237038P.
PR    02-OCT-2000; 2000US-0237039P.
PR    02-OCT-2000; 2000US-0237040P.
PR    13-OCT-2000; 2000US-0239935P.
PR    20-OCT-2000; 2000US-0240960P.
PR    20-OCT-2000; 2000US-0241785P.
PR    20-OCT-2000; 2000US-0241809P.
PR    01-NOV-2000; 2000US-0244617P.
PR    17-NOV-2000; 2000US-0249299P.
PR    08-DEC-2000; 2000US-0251856P.
PR    08-DEC-2000; 2000US-0251868P.
PR    08-DEC-2000; 2000US-0251869P.
```

```
XX
PA    (ROSE/)  ROSEN C A.
PA    (RUBE/)  RUBEN S M.
PA    (BARA/)  BARASH S C.
XX
PI    Rosen CA,  Ruben SM,  Barash SC;
XX
DR    WPI; 2003-147444/14.
DR    P-PSDB; ABU54924.
XX
PT    New polypeptides and nucleic acids, useful in gene therapy for treating,
PT    inhibiting or preventing e.g. neural, immune system, muscular,
PT    respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT    renal disorders.
XX
PS    Claim 1; SEQ ID NO 22; 402pp; English.
XX
CC    The invention relates to human novel polypeptides and their associated
CC    polynucleotides. The polypeptides and polynucleotides are useful in gene
CC    therapy for treating, inhibiting or preventing neural disorders, immune
CC    system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC    and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC    nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC    gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC    (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC    heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC    renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC    leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC    appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC    related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC    infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC    human novel polynucleotides of the invention
XX
SQ    Sequence 1311 BP; 306 A; 359 C; 399 G; 239 T; 0 U; 8 Other;

Query Match             35.3%;  Score 945;  DB 7;  Length 1311;
Best Local Similarity   81.1%;  Pred. No. 3.5e-217;
Matches 1267;  Conservative    6;  Mismatches    7;  Indels  282;  Gaps    4;
```

```
Qy      56  TCCGGGTTCGGCGCGGGGCGGGGATGTGAATCCCGATGGAGGCCGCCGAGGAAGGCAAGC  115
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5  TCCGGGTTCGGCGCGGGGCGGGGAATCCCGATGGAGCGGAATCCCGATGGAGGCCGCCGAGGAAGGCAAGC   64

Qy     116  AGTCGCCGCCGCCCGAGCCTGGGACGGCTCCTGCGTCTGGGCGCGGAGGAGGGCGAGC  175
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db      65  AGTCGCCGCCGCCCGCAGCCCTGGGGACGGGCTCCTGCGTCTGGGCGCGGAGGAGGGCGAGC  124

Qy     176  CGCACGTCCTCCTGAGGAAGCGGAGTGGACCATCGGGCGGAGACCAGAGGTTGCGACCTTT  235
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125  CGCACGTCCTCCTGAGGAAGCGGAGTGGACCATCGGGCGGAGACGAGAGGTTGCGACCTTT  184

Qy     236  CCTTCCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAAT  295
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db     185  CCTTCCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAAT  244

Qy     296  CAGGTCAGGTGACACTGGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGG  355
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     245  CAGGTCAGGTGACACTGGAAGATACCAGCACCAGTGGAACAGTGATTAACAARCTGAARG  304

Qy     356  TTGTTAAGAAGGCAGACATGCCCTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGA  415
                ||||||||:|||||||||||||||||||||||||||||||||||||||
Db     305  TTGTTAAGAARCAGACATGCCCTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGA  364

Qy     416  AGAATGAACCGGAACAACAAGTGGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGGCA  475
                |||||||||||||||
Db     365  AGAATGAACCGGAACACA-------------------------------------------  382

Qy     476  TGACACAAGAATCCTTTGAAGCTAACAAGGAAAAATTGTTCCATGGGACCAAAGATACCT  535
Db     383  -----------------------------------------------------------  382

Qy     536  CAGGTGCAGGTGCAGGGCGAGGGCCGATCCCGGGTCCCTCCGTCGCCCGCCACTC       595
```

```
PT  diagnosing, preventing, treating or ameliorating medical conditions and
PT  used as food additives or preservatives.
XX
PS  Claim 1; SEQ ID NO 22; 980pp; English.
XX
CC  The invention relates to isolated nucleic acid molecules and their
CC  encoded secreted proteins. The nucleic acids and proteins are used to
CC  prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC  rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC  in diagnosing a pathological condition or susceptibility to a
CC  pathological condition. Antibodies to the proteins can also be used in
CC  alleviating symptoms associated with the disorders and in diagnostic
CC  immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC  (ELISA). Disorders which are diagnosed or treated include autoimmune
CC  diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC  neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC  arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC  nervous system disorders e.g. Alzheimer's disease, infections caused by
CC  bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC  and many other disorders listed in the specification. The polypeptides
CC  can also be used to aid wound healing and epithelial cell proliferation,
CC  to prevent skin aging due to sunburn, to maintain organs before
CC  transplantation, for supporting cell culture of primary tissues, to
CC  regenerate tissues and in chemotaxis. The polypeptides can also be used
CC  as a food additive or preservative to increase or decrease storage
CC  capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC  minerals, cofactors and other nutritional components. The present
CC  sequence encodes a novel secreted protein of the invention. Note: The
CC  sequence data for this patent did not form part of the printed
```

```
RESULT 8
ABX73184
ID   ABX73184 standard; DNA; 1311 BP.
```

```
Query Match      35.3%;  Score 945;  DB 4;  Length 1311;
Best Local Similarity  81.1%;  Pred. No. 3.5e-217;
Matches 1267;  Conservative    6;  Mismatches    6;  Indels  282;  Gaps   4;
```

```
Qy   56  TCCGGGTTCGGCGCGCCGGCGGGGGGGGGATGGAATCCCGATGGAGCGGCCCGAGGAAGGCAAGC  115
         |||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||||||
Db    5  TCCGGGTTCGGCGCGCCGGCGGGGGGGGGATGGAATCCCGATGGAGCGGCCCGAGGAAGGCAAGC   64

Qy  116  AGTCGCCGCCGCCGCCGCCGCCGCCCTGGGACGGCTCCTGCGTCTTCGGGCCGCTCCTGCGTCTTCGGGCCGGAGGGCGAGC  175
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   65  AGTCGCCGCCGCCGCCGCCGCCGCCCTGGGACGGCTCCTGCGTCTTCGGGCCGCTCCTGCGTCTTCGGGCCGGAGGGCGAGC  124

Qy  176  CGCACGTCCTCTGAGGAAGCGGAGTGGACCATCGGCGGAGACCAGGTTGCGACCTTT  235
         ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db  125  CGCACGTCCTCTGAGGAAGCGGAGTGGACCATCGGCGGAGACCAGGTTGCGACCTTT  184

Qy  236  CCTTCCCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAAT  295
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  185  CCTTCCCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAAT  244

Qy  296  CAGGTCAGGTGACACTGGAAGATACCAGCACCACGTGATGGAGAACAGTGATTAACAGCTGAAGG  355
         |||||||||||||||||||||||||||||||||:|||:||||||||||||||:||||
Db  245  CAGGTCAGGTGACACTGGAAGATACCAGCACCACGTGGGAACAGTGATTAACAARCTGAARG  304

Qy  356  TTGTTAAGAAGCAGACATGCCCTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGA  415
         ||||||||:|||||||||||||||:|||||||||||||||||||||||||||||||||
Db  305  TTGTTAAGAARCAGACATGCCCTTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGA  364

Qy  416  AGAATGAACCGGAACAACAACGTGGCATACCTACCTGCCATACCTCTTATGGAAGTCTTTAAGTGAAAAAGCAAGGCA  475
         |||||||||||||||
Db  365  AGAATGAACCGGAACACA  382

Qy  476  TGACACAAGAGAATCCTTTGAAGCTAACAAGGAAAATGTGTTCATGGGACCAAAGATACCT  535
Db  383  --------------  382

Qy  536  CAGGTGCAGGTGCAGGGGCGAGGGGCCGATCCCCGGGTCCCGTCGGTCGCCCGGCCACTC  595
Db  383  --------------  382

Qy  596  AGGTGTGCTTTGAGGAACCACAGCCACCAGCCAATCAACATCGACGTCAGACCTCTTCCCCACAGCCT  655
Db  383  --------------  382
```

```
Qy  656  CGGGCCTCTTCCACGGAGGAGCCTTCCTCCTGCGAGGCGGAGAGCGTTCCTCCCAGTTGTGGGCTG  715
         ||||||
Db  383  ---------------------------------------------------------GGTCTG  388

Qy  716  GGGGTGGTGGCATCTCCCTAAAGGAAGGAGTGGTCCCTCTGTGGCAAGTGATGAAGTCTCCA  775
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  389  GGGGTGGTGGCATCTCCCTAAAGGAAGGAGTGGTCCCTCTGTGGCAAGTGATGAAGTCTCCA  448

Qy  776  GCTTTGCCTCAGCTCTCCAGACAGCAGAAGACTCGCTCCTTTCGTCGTTGGAACCCCAGG  835
         |||||||||||||||||||||||||||||||||||:|||||:||||||||||||||||||
Db  449  GCTTTGCCTCAGCTCTCCAGACAGCAGAAGACTCGGTCCTTTTCGTCGTTGGAACCCCAGG  508

Qy  836  ATCAGGAGGATTTGGAGCCCGTGAGAGAAAATGAGGAGGAGAGGGGGACCTTGACCTGA  895
         |||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db  509  ATCAGGAGGATTTGGAGCCCGTGAAGAAGAAAATGAGGAGGAGAGGGGGACCTTGACCTGA  568

Qy  896  ACGGGCAGTTGTTGGTCGCACACCGCGTAGAAATGCCCAAACCGTCCACGAGAGACGTCA  955
         |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db  569  ACGGGCAGTTGTTGGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAGAGACGTCA  628

Qy  956  GAGCAGCGGCTGGGAAGCCAGACAGATGGAGGAGACCGTGACATGCATCATCCTGCCAGG  1015
         |||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db  629  GAGCAGCGGCTGGGAAGCCAGACAAGATGGAGGAGACCGTGACATGCATCATCCTGCCAGG  688

Qy 1016  ACCTGCTGCACGACTGCGTGAGTTGCAGCCCTGCATGCACCACGTTCTGCGCGGCTTGCT  1075
         |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db  689  ACCTGCTGCACGACTGCGTGAGTTTGCAGCCCTGCATGCACCACGTTCTGCGCGGCTTGCT  748

Qy 1076  ACTCGGCTGGATGAGCGGCTCGTCCCTGTGTCCCTGTGTCCTCTACCTGCCCGTGTCCCGTGGAGCGGA  1135
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  749  ACTCGGCTGGATGAGCGGCTCGTCCCTGTGTCCCTGTGTCCTCTACCTGCCCGTGTCCCGTGGAGCGGA  808

Qy 1136  TCTGTAAAAACCACATCCTCAACAACCTGTGGAAGCATACCTCATCCAGCATCCAGACA  1195
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  809  TCTGTAAAAACCACATCCTCAACAACCTGTGGAAGCATACCTCATCCAGCATCCAGACA  868

Qy 1196  AGAGTCGCAGTGAAGAAGATGTCAAAGTATGATGCCAGGAATAAAATCACTCAAGACA  1255
         ||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db  869  AGAGTCGCAGTGAAGAAGATGTGAAAGTATGATGCCAGGAATAAAATCACTCAAGACA  928

Qy 1256  TGCTGCAGCCCAAAGTCAGGCGGCTCTTTTCTGATGAAGAAGGAGTTCAGAGGACCTTT  1315
         ||||||||||||||||||||||||||||||:||||||||||||||||||||||||:||
Db  929  TGCTGCAGCCCAAAGTCAGGCGGCTCTTTTTCTGATGAAGAAGGAGTTCAGAGGACCTGC  988

Qy 1316  TGGAGCTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATTAGCCAGTGTGCC  1375
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  989  TGGAGCTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATTAGCCAGTGTGCC  1048

Qy 1376  GGCAGTGTCCTGAGTACAGAAGGCAGCGGCGGGGGGGATGGCACCCGAGG  1435
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1049  GGCAGTGTCCTGAGTACAGAAGGCAGCGGCGGGGGGGATGGCACCCGAGG  1108

Qy 1436  GCGAGCCAGAGAGCCCACAGAGCCCCTGGGGGATGCAAGGAAGCCACGCCCTGTGACGA  1495
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1109  GCGAGCCAGAGAGCCCACAGAGCCCCTGGGGGATGCAAGGAAGCCACGCCCTGTGACGA  1168

Qy 1496  CAGAGTCCAGGATTACGTTGCCCTCTGCAAGGAAGCCACGCCCTGTGCACCTGCT  1555
         ||||   |||||||||||:|||||||||||||||||||||||||||||||||||||
Db 1169  CAGT---CAGGATTACGTGTGCCCTCTGCAAGGAAGCCACGCCCTGTGCACCTGCT  1224

Qy 1556  TCCAGCCCATGCCCGACCCGGAGGCGGAGGCCGGACGGAGGACAGGA-CCCGGCGGTGTCGCCCCTCAG  1614
         |:|||||||||||||||||||||||||||||||||||||||||||||:||||:||||:|:||||||
Db 1225  T-CAGCCCATGCCCGACCCGGAGGCGGAGGCCGGACGGAGGACAGGCAACAGGACCCCGGCGTGKCCGGSCCCTNAA  1283

Qy 1615  CA  1616
         ||
Db 1284  CA  1285
```

```
XX
PR   31-JAN-2000;  2000US-0179065P.
PR   04-FEB-2000;  2000US-0180628P.
PR   24-FEB-2000;  2000US-0184664P.
PR   02-MAR-2000;  2000US-0186150P.
PR   16-MAR-2000;  2000US-0189874P.
PR   17-MAR-2000;  2000US-0190076P.
PR   18-APR-2000;  2000US-0198123P.
PR   19-MAY-2000;  2000US-0205515P.
PR   07-JUN-2000;  2000US-0209467P.
PR   28-JUN-2000;  2000US-0214486P.
PR   30-JUN-2000;  2000US-0215135P.
PR   07-JUL-2000;  2000US-0216647P.
PR   07-JUL-2000;  2000US-0216880P.
PR   11-JUL-2000;  2000US-0217487P.
PR   11-JUL-2000;  2000US-0217496P.
PR   14-JUL-2000;  2000US-0218290P.
PR   26-JUL-2000;  2000US-0220963P.
PR   26-JUL-2000;  2000US-0220964P.
PR   14-AUG-2000;  2000US-0224518P.
PR   14-AUG-2000;  2000US-0224519P.
PR   14-AUG-2000;  2000US-0225213P.
PR   14-AUG-2000;  2000US-0225214P.
PR   14-AUG-2000;  2000US-0225266P.
PR   14-AUG-2000;  2000US-0225267P.
PR   14-AUG-2000;  2000US-0225268P.
PR   14-AUG-2000;  2000US-0225270P.
PR   14-AUG-2000;  2000US-0225447P.
PR   14-AUG-2000;  2000US-0225757P.
PR   14-AUG-2000;  2000US-0225758P.
PR   14-AUG-2000;  2000US-0225759P.
PR   18-AUG-2000;  2000US-0226279P.
PR   22-AUG-2000;  2000US-0226681P.
PR   22-AUG-2000;  2000US-0226868P.
PR   22-AUG-2000;  2000US-0227182P.
PR   23-AUG-2000;  2000US-0227009P.
PR   30-AUG-2000;  2000US-0228924P.
PR   01-SEP-2000;  2000US-0229287P.
PR   01-SEP-2000;  2000US-0229343P.
PR   01-SEP-2000;  2000US-0229344P.
PR   01-SEP-2000;  2000US-0229345P.
PR   05-SEP-2000;  2000US-0229509P.
PR   05-SEP-2000;  2000US-0229513P.
PR   06-SEP-2000;  2000US-0230437P.
PR   06-SEP-2000;  2000US-0230438P.
PR   08-SEP-2000;  2000US-0231242P.
PR   08-SEP-2000;  2000US-0231243P.
PR   08-SEP-2000;  2000US-0231244P.
PR   08-SEP-2000;  2000US-0231413P.
PR   08-SEP-2000;  2000US-0231414P.
PR   08-SEP-2000;  2000US-0232080P.
PR   08-SEP-2000;  2000US-0232081P.
PR   12-SEP-2000;  2000US-0219868P.
PR   14-SEP-2000;  2000US-0232397P.
PR   14-SEP-2000;  2000US-0232398P.
PR   14-SEP-2000;  2000US-0232399P.
PR   14-SEP-2000;  2000US-0232400P.
PR   14-SEP-2000;  2000US-0232401P.
PR   14-SEP-2000;  2000US-0233063P.
PR   14-SEP-2000;  2000US-0233064P.
PR   14-SEP-2000;  2000US-0233065P.
PR   21-SEP-2000;  2000US-0234223P.
PR   21-SEP-2000;  2000US-0234274P.
PR   25-SEP-2000;  2000US-0234997P.
PR   25-SEP-2000;  2000US-0234998P.
PR   26-SEP-2000;  2000US-0235484P.
PR   27-SEP-2000;  2000US-0235834P.
PR   27-SEP-2000;  2000US-0235836P.
PR   29-SEP-2000;  2000US-0236327P.
PR   29-SEP-2000;  2000US-0236367P.
PR   29-SEP-2000;  2000US-0236368P.
PR   29-SEP-2000;  2000US-0236369P.
PR   29-SEP-2000;  2000US-0236370P.

PR   02-OCT-2000;  2000US-0236802P.
PR   02-OCT-2000;  2000US-0237037P.
PR   02-OCT-2000;  2000US-0237038P.
PR   02-OCT-2000;  2000US-0237039P.
PR   02-OCT-2000;  2000US-0237040P.
PR   13-OCT-2000;  2000US-0239935P.
PR   13-OCT-2000;  2000US-0239937P.
PR   20-OCT-2000;  2000US-0240960P.
PR   20-OCT-2000;  2000US-0241221P.
PR   20-OCT-2000;  2000US-0241785P.
PR   20-OCT-2000;  2000US-0241786P.
PR   20-OCT-2000;  2000US-0241787P.
PR   20-OCT-2000;  2000US-0241808P.
PR   20-OCT-2000;  2000US-0241809P.
PR   20-OCT-2000;  2000US-0241826P.
PR   01-NOV-2000;  2000US-0244617P.
PR   08-NOV-2000;  2000US-0246474P.
PR   08-NOV-2000;  2000US-0246475P.
PR   08-NOV-2000;  2000US-0246476P.
PR   08-NOV-2000;  2000US-0246477P.
PR   08-NOV-2000;  2000US-0246478P.
PR   08-NOV-2000;  2000US-0246523P.
PR   08-NOV-2000;  2000US-0246524P.
PR   08-NOV-2000;  2000US-0246525P.
PR   08-NOV-2000;  2000US-0246526P.
PR   08-NOV-2000;  2000US-0246527P.
PR   08-NOV-2000;  2000US-0246528P.
PR   08-NOV-2000;  2000US-0246532P.
PR   08-NOV-2000;  2000US-0246609P.
PR   08-NOV-2000;  2000US-0246610P.
PR   08-NOV-2000;  2000US-0246611P.
PR   08-NOV-2000;  2000US-0246613P.
PR   17-NOV-2000;  2000US-0249207P.
PR   17-NOV-2000;  2000US-0249208P.
PR   17-NOV-2000;  2000US-0249209P.
PR   17-NOV-2000;  2000US-0249210P.
PR   17-NOV-2000;  2000US-0249211P.
PR   17-NOV-2000;  2000US-0249212P.
PR   17-NOV-2000;  2000US-0249213P.
PR   17-NOV-2000;  2000US-0249214P.
PR   17-NOV-2000;  2000US-0249215P.
PR   17-NOV-2000;  2000US-0249216P.
PR   17-NOV-2000;  2000US-0249217P.
PR   17-NOV-2000;  2000US-0249218P.
PR   17-NOV-2000;  2000US-0249244P.
PR   17-NOV-2000;  2000US-0249245P.
PR   17-NOV-2000;  2000US-0249264P.
PR   17-NOV-2000;  2000US-0249265P.
PR   17-NOV-2000;  2000US-0249297P.
PR   17-NOV-2000;  2000US-0249299P.
PR   17-NOV-2000;  2000US-0249300P.
PR   01-DEC-2000;  2000US-0250160P.
PR   01-DEC-2000;  2000US-0250391P.
PR   05-DEC-2000;  2000US-0251030P.
PR   05-DEC-2000;  2000US-0251988P.
PR   05-DEC-2000;  2000US-0256719P.
PR   06-DEC-2000;  2000US-0251479P.
PR   08-DEC-2000;  2000US-0251856P.
PR   08-DEC-2000;  2000US-0251868P.
PR   08-DEC-2000;  2000US-0251869P.
PR   08-DEC-2000;  2000US-0251989P.
PR   08-DEC-2000;  2000US-0251990P.
PR   11-DEC-2000;  2000US-0254097P.
PR   05-JAN-2001;  2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   P-PSDB; AAU15856.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
```

```
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO08972.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT   and treating e.g. leukemia, inflammation and immune disorders.
XX
PS   Claim 1; SEQ ID NO 8961; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2259 BP; 531 A; 612 C; 615 G; 501 T; 0 U; 0 Other;

Query Match        48.3%;  Score 1294.2;  DB 4;  Length 2259;
Best Local Similarity   99.6%;  Pred. No. 3.3e-301;
Matches 1308;  Conservative     0;  Mismatches     3;  Indels     2;  Gaps     1;
```

Qy	1354 AGCCAGCCATACGTCGTGTGCCGCAGTGTCCTGAGTACAGAAGGCAGGCGGCGGCAGCCT	1413
Db	1965 AGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCGGCGGCAGCCT	1906
Qy	1414 CCCACTGCCCAGCACCCGAGGGGCGAGCAGCCAGGCCCCTGGGGGATGCACCC	1473
Db	1905 CCCCACTGCCCAGCACCCGAGGGGCGAGCAGCCAGGCCCCTGGGGATGCACCC	1846
Qy	1474 TCCAGTCGTCGTCAGCCTGACGACAGCAGTCAGGATTACGTGCCCTCTGCAAGGAAGC	1533
Db	1845 TCCAGTCGTCCGTCAGCCTGACGACAGCAGTCAGGATTACGTGCCCTCTGCAAGGAAGC	1786
Qy	1534 CACGCCCTGCACCTGCTTCCAGCCCATGCCCGACCGGCGGAGCCGAGCAG	1593
Db	1785 CACGCCCTGTGCACCTGCTTCCAGCCCATGCCCGACCGGCGGAGCCGAGCAG	1726
Qy	1594 GACCCGGCGGTCGCCCTCAGCAGTGTGCGGTCCTGCGGTGTGCCGCCTTCTGCCACCTGTAC	1653
Db	1725 GACCCGGCGTGTCGCCCCTCAGCAGTGTGCGGTCCTGCGGTGTGCCGCCTTCTGCCACCTGTAC	1666
Qy	1654 TGGGGCTGCACCCGGACCCGGACCGGCGTGCTACGGCGTGCTGCCCTGGCCCTG...CGTTTGTGAGCTCAACCTG	1713
Db	1665 TGGGGCTGCACCCGGACCCGGACCGGCGTGCTACGGCGTGCTGCCCTGGCCCTGGCGTTTGTGAGCTCAACCTG	1606
Qy	1714 GGTGACAAGTGTCTGGACGGCGTGCTGAACAACAACAGTCAGAGTCAGACATCCTGAAG	1773
Db	1605 GGTGACAAGTGTCTGGACGGCGTGCTGAACAACAACAGTCAGAGTCAGACATCCTGAAG	1546
Qy	1774 AATTACCTGCAACCAGAGGTTTGACATGAAAAAACATGTTGACCGAGAGCCTCTGGCT	1833
Db	1545 AATTACCTGCAACCAGAGGTTTGACATGAAAAAACATGTTGACCGAGAGCCTCTGGCT	1486
Qy	1834 CTCCAGCGGGGAGTGTTTCTGCTTGCTGATTACAGAGTCACGGGAGACACCGTTCTGTGT	1893
Db	1485 CTCCAGCGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGT	1426
Qy	1894 TACTGCTGTGGCCTGGCCAGTCGCCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCT	1953
Db	1425 TACTGCTGTGGCCTGGCCAGTGCGCCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCT	1366
Qy	1954 GCTTCCGAGTTGCCAGTGGCCTACTGGCCCGTAACATCCCGTCCTGACTGCTACTGGGGCCGTAACTGC	2013
Db	1365 GCTTCCGAGTTGCCAGTGGCCTACTGGCCCGTAACATCCCGTCCTGACTGCTACTGGGGCCGTAACTGC	1306
Qy	2014 CGCACTCAGGTGAAAGCTCACCACCGCCATGAAATTCAATCATATCTGTGAACAGACAAGG	2073
Db	1305 CGCACTCAGGTGAAAGCTCACCACCGCCATGAAATTCAATCATATCTGTGAACAGACAAGG	1246
Qy	2074 TTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGAGGTGAAGAGAGCG	2133
Db	1245 TTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGAGGTGAAGAGAGCG	1186
Qy	2134 TGTTTTAAAATACAGAGACAAGCACGTCAAGTGTTTCACAGCCCCCTGAGGGAAGGG	2193
Db	1185 TGTTTTTAAAATACAGAGACAAGCACGTCAAGTGTTTCACAGCCCCCTGAGGGGAAGGG	1126
Qy	2194 ACGCAGGGTCTCCGACAGGTGCTCTGGGTGACTCTTCTGTGGAGTTTTTACCCTCTGA	2253
Db	1125 ACGCAGGGTCTCCGACACAGGTGCTCTGGGGTGACTCTTCTGTGTGGAGC--TTTACCCTCTGA	1068
Qy	2254 GTGAGACCCTCCCAGACCCCGGGGGGCGCCAGCCCGCCCTTCTGTGAGCCTGGGCAG	2313
Db	1067 GTGAGACCCTCCCAGAGCCCCGGGGGGGCGCCAGCCCGCCCTCCTGGGTGAGCCGTGGGCAG	1008
Qy	2314 GGCTCGTGTGGTGGCATCAGCAGGACGAAGCCCTTTCTGTAACATGCGGCCGTCCCGCC	2373
Db	1007 GGCTCGTGTGGTGGCATCAGCACCAGGACGAAGCCCTTTCTGTAACATGCGGCCGTCCCGCC	948
Qy	2374 GAGAGGGGCAGTTTTGCTCTTTGTACATTTTCGGAAACTACAGTTAAAGCAAGAAGTCTG	2433
Db	947 GAGAGGGGCAGTTTTGCTCTTTTGTACATTTTCGGAAACTACAGTTAAAGCCGGAAGTCTG	888
Qy	2434 TTTTCAGGAAAAGTTCAAGGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAA	2493
Db	887 TTTTCAGGAAAAGTTCAAGGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAA	828
Qy	2494 GGGAGCATAAGTTACAGCCTACAGGACGTACAACAATATCCTGCTGCTGGGAAAACCACA	2553
Db	827 GGGAGCATAAGTTACAGCCTACAGGACGTACAACAATATCCTGCTGCTGGGAAAACCACA	768
Qy	2554 GCATTTATCTATTTTTTATTTTAATTAGGTTGGTGCTTATCTTCTAATAAGATTTAAAT	2613
Db	767 GCATTTATCTATTTTTTATTTTAATTAGGTTGGTGCTTATCTTCTAATAAGATTTAAAT	708
Qy	2614 GTCACAAACTGTAGCACAAATAATAATAATTATAATTTACAAATTGACAAAAA	2666
Db	707 GTCACAAACTGTAGCACAAATAATAATAATTATAATTTACAAATTGACTAAAA	655

```
RESULT 7
AAS25843
ID   AAS25843 standard; cDNA; 1311 BP.
XX
AC   AAS25843;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human cDNA encoding a novel secreted protein, Seq ID 22.
XX
KW   Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
```

```
Qy   391 GATGTCATCTACTTGGTGTACAGGAAGAATGAACCGGAACAACACAAGCTGGCATACCTCTAT 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GATGTCATCTACTTGGTGTACAGGAAGAATGAACCGGAACAACACAAGCTGGCATACCTCTAT 360

Qy   451 GAATCTTTAAGTGAAAAGCAAGGCATGACACAAGAATCCTTTGAAGCTAACAAGGAAAAT 510
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GAATCTTTAAGTGAAAAGCAAGGCATGACACAAGAATCCTTTGAAGCTAACAAGGAAAAT 420

Qy   511 GTGTTCCATGGGACCAAAGATACCTCAGGTGCAGGTGCAGGGGCCGATCCCCGG 570
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTGTTCCATGGGACCAAAGATACCTCAGGTGCAGGTGCAGGGGCCGAGGGGCCGATCCCCGG 480

Qy   571 GTCCCCGTCGTCGCCCGCCACTCAGGTGCGTGCTTTGAGGAACCACCAGCCATCAACATCG 630
         ||||| ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   481 GTCCCTTCCGTCGTCGCCCGCCACTCAGGTGTGCTTTGAGGAACCACCAGCCATCAACATCG 540

Qy   631 ACGTCAGACCCTCTTCCCCACAGCCTCGGCGCCTTCTTCCACGAGCCTTCCTCGCAGGGCGA 690
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 ACGTCAGACCCTCTTCCCCACAGCCTCGGCGCCTTCTTCCACGAGCCTTCCTCGCAGGGCGA 600

Qy   691 GAGCGGTTCCTCCAGTGTGTGGGTTCGGGGGTGGTGGCATCTCCCCTAAAGGAAGTGTCCC 750
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GAGCGGTTCCTCCAGTGTGTGGGTTCGGGGGTGGTGGCATCTCCCCTAAAGGAAGTGTCCC 660

Qy   751 TCTGTGCGCAAGTGATGAAGTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGCAGAAGACTGCG 810
         |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TCTGTGCGGCAAGTGATGAAGTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGCAGAAGACTGCG 720

Qy   811 TCCTTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAAATG 870
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 TCCTTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAAATG 780

Qy   871 AGAGGAGATGGGAGACCTTGACCTGACGCGGGACCAGTTGTTGTCGCACAACCGCGTAGAAAT 930
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 AGAGGAGATGGGAGACCTTGACCTGACGCGGGACCAGTTGTTGTCGCACAACCGCGTAGAAAT 840

Qy   931 GCCCAAACCGTCCACGAGGACTCAGAGCAGCGGCTGGGAAGCCAGCAAGATGGAGGAG 990
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 GCCCAAACCGTCCACGAGGACTCAGAGCAGCGGCTGGGAAGCCAGCAAGATGGAGGAG 900

Qy   991 ACGCTGACACATGCATCATCTGCGCGGGACCTGCTGCTGCACGACTGCGTGAGTTGCAGCCCTGC 1050
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 ACGCTGACACATGCATCATCTGCGCGGGACCTGCTGCTGCACGACTGCGTGAGTTGCAGCCCTGC 960

Qy  1051 ATGCACACGTTCTGCGCGGGCTTGCTACTCGGCGGATGGAGCGGCTCGTGCCCTGTGTCCT 1110
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 ATGCACACGTTCTGCGCGGGCTTGCTACTCGGCGGATGGAGCGGCTCGTGCCCTGTGTCCT 1020

Qy  1111 ACCTGCCGCGCTGTCCCGGAGCGGGATCTGTAAAAACCACATTCCTCAACACACCTCGGTGGAA 1170
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ACCTGCCGCGCTGTCCCGGAGCGGGATCTGTAAAAACCACATTCCTCAACACACCTCGGTGGAA 1080

Qy  1171 GCATACCTCATCCAGCATCCAGACAGAGTCGCAGTGAAGATGTGAAGAAGATGTGCAAAGTATGGAT 1230
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GCATACCTCATCCAGCATCCAGACAGAGTCGCAGTGAAGATGTGAAGAAGATGTGCAAAGTATGGAT 1140

Qy  1231 GCCAGGAATAAAATCACTCAAGACATGCTGAGCGCCCAAAGTCAGGGGTCTTTTTCTGAT 1290
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 GCCAGGAATAAAATCACTCAAGACATGCTGAGCGCCCAAAGTCAGGGGTCTTTTTCTGAT 1200

Qy  1291 GAAGAAGGGAGTTCAGAGGACCTGCTGGAGCGTGTCAGACGTTGACAGTGAGTCCTCAGAC 1350
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 GAAGAAGGGAGTTCAGAGGACCTGCTGGAGCGTGTCAGACGTTGACAGTGAGTCCTCAGAC 1260

Qy  1351 ATTAGCCAGCCATACGTCGTGTGCCCGGCAGTGTCCTGAGTACAGAGGGCAGGCGGCGCAG 1410
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 ATTAGCCAGCCATACGTCGTGTGCCCGGCAGTGTCCTGAGTACAGAGGGCAGGCGGCGCAG 1320

Qy  1411 CCTCCCACTGCCCCAGCACCCCAGGCGGCCCCAGAGGCCCACAGGCCCTGGGGGATGCA 1470
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 CCTCCCACTGCCCCAGCACCCCAGGCGGCCCCAGAGGCCCACAGGCCCTGGGGGATGCA 1380

Qy  1471 CCCTCCACGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTGTGCCCTCTGTGCCAAGGA 1530
         ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  1381 CCCTCCACGTCCGTCAGCCTGACGACGACGACGACCAGGATTACGTGTGCCCTCTGTGCCAAGGA 1440

Qy  1531 AGCCACGCCGTGACCTGCTGCTGCCTGCTTCAGCCCATGCCCATGCCCGACCCGGAGAGCGGAGCGCGAG 1590
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 AGCCACGCCCTGTCACCTGCTGCTGCCTGCTTCAGCCCATGCCCATGCCCGACCCGGAGAGCGGAGCGCGAG 1500

Qy  1591 CAGGACCCGGCGTGTCGCCCCTCAGCAGTGTGCGGTCTGCGTGCCTTTCTGCCACCTG 1650
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 CAGGACCCGGCGTGTCGCCCCTCAGCAGTGTGCGGTCTGCGTGCCTTTCTGCCACCTG 1560

Qy  1651 TACTGGGGCTGCACCCGGACCGGCTGCTACGGCCCCGTTTTTGTGAGCTCAAC 1710
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 TACTGGGGCTGCACCCGGACCGGCTGCTACGGCCCCGTTTTTGTGAGCTCAAC 1620

Qy  1711 CTGGGTGACAAGTGTCTGGACGGCGGCGTGCTGAACAACAGCTACGAGTCAGATCCTG 1770
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 CTGGGTGACAAGTGTCTGGACGGCGGCGTGCTGAACAACAGCTACGAGTCAGATCCTG 1680

Qy  1771 AAGAATTACCTGGCAACCAGAGAGGTTTGACATGGAAAAACATGTTGACCGAGAGCCTCGTG 1830
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 AAGAATTACCTGGCAACCAGAGAGGTTTGACATGGAAAAACATGTTGACCGAGAGCCTCGTG 1740

Qy  1831 GCTCTTCCAGCAGGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTG 1890
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 GCTCTTCCAGCAGGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTG 1800

Qy  1891 TGTTACTGCGTGTGGGCCTGCGCAGCTTCCGTGAGCTGACCTATCGGCAGAACATT 1950
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 TGTTACTGCGTGTGGGCCTGCGCAGCTTCCGTGAGCTGACCTATCGGCAGAACATT 1860

Qy  1951 CCTGCTTCCCAGTTGCCAGTGGCCGTAAACATCCCGTCCTGACTGCTACTGGGGCCGTAAC 2010
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 CCTGCTTCCCAGTTGCCAGTGGCCGTAAACATCCCGTCCTGACTGCTACTGGGGCCGTAAC 1920

Qy  2011 TGCCCGCACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATCATATCGTGACACAGACA 2070
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 TGCCCGCACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATCATATCGTGACACAGACA 1980

Qy  2071 AGGTTCAAAAACTAA 2085
         |||||||||||||||
Db  1981 AGGTTCAAAAACTAA 1995
```

RESULT 5
AAF89709
ID   AAF89709 standard; DNA; 1995 BP.
XX
AC   AAF89709;
XX
DT   22-AUG-2001 (first entry)
XX
DE   Nucleotide sequence of a human ring finger protein designated FHAR1.
XX
KW   FHAR1; RING finger protein; cancer; vaccine; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1995
FT   /tag= a
FT   /product= "ring finger protein FHAR1"
XX
PN   WO200142430-A1.
XX
PD   14-JUN-2001.
XX
PF   07-DEC-2000; 2000WO-US033094.
XX
PR   08-DEC-1999; 99US-00456876.
XX
PA   (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI   Zhou B,   Zhu Y,   Chaturvedi P,   Hurle MR,   Li X;
XX
DR   WPI; 2001-381663/40.
DR   P-PSDB; AAB83843.
XX
PT   New FHAR1 polypeptide, a member of the RING finger protein family for
PT   diagnosing and treating cancer, and for use in anti-cancer vaccines.
XX
PS   Claim 2; Page 18-19; 28pp; English.
XX
CC   The present sequence encodes a FHAR1 polypeptide, which is a member of
CC   the RING finger protein family. FHAR1 is useful in the treatment of
CC   cancer, and as a vaccine for inducing an immunological response in a
CC   mammal. FHAR1 polynucleotides may also be used as a diagnostic reagent
CC   through detection of mutations in the associated gene, and for chromosome
CC   localization studies, and tissue expression studies. FHAR1 antibodies are
CC   useful to isolate and to identify clones expressing the polypeptides, or
CC   to purify the polypeptides by affinity chromatography and to treat cancer
XX
SQ   Sequence 1995 BP; 465 A; 563 C; 580 G; 387 T; 0 U; 0 Other;

Query Match   74.5%; Score 1995; DB 5; Length 1995;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   91 ATGGAGCGGCCGCCGGAGGAAGGCAAGCAGTCGCCGCCGCCGGGGACCGGCTCCTG 150
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGAGCGGCCGCCGGAGGAAGGCAAGCAGTCGCCGCCGCCGGGGACCGGCTCCTG 60

Qy   151 CGTCTGGGCCGCGGAGGAGGGGCGAGCCGCGCACGTCCTCCTCCTGAGGAGTGGACCATC 210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CGTCTGGGCCGCGGAGGAGGGGCGAGCCGCGCACGTCCTCCTCCTGAGGAGTGGACCATC 120

Qy   211 GGGGCCGGAGACGAGGTTGCGACCTTTCCTTCCCAGCAATAAACTGGTCTCTGGAGATCAC 270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGGGCCGGAGACGAGGTTGCGACCTTTCCTTCCCAGCAATAAACTGGTCTCTGGAGATCAC 180

Qy   271 TGTAGAATTGTAGTGGATGGATGAAAAATCAGGTGACACTGGAGAAATACCAGCACCAGT 330
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGTAGAATTGTAGTGGATGGATGAAAAATCAGGTGACACTGGAGAAATACCAGCACCAGT 240

Qy   331 GGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGCAGCATGCCCTTTACAGACTGGG 390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGCAGCATGCCCTTTACAGACTGGG 300

Qy   1636 CCTTTCTGCCACCTGTACTGGGGCGTGCACCGGACCGGCGTGCTACGGCCGGCTGCCTGGCCCCG 1695
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1518 CCTTTCTGCCACCTGTACTGGGGCGTGCACCGGACCGGCGTGCTACGGCCGGCTGCCTGGCCCCG 1577

Qy   1696 TTTTGTGAGCTCAACCTGGGTGACAAGTGTCTGGACGGCGTGCTGAACAACACAGCTAC 1755
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1578 TTTTGTGAGCTCAACCTGGGTGACAAGTGTCTGGACGGCGTGCTGAACAACACAGCTAC 1637

Qy   1756 GAGTCAGACATCCTGAAGAATTACCTGGCAACCAGAGGTTGACATGGAAAAACATGTTG 1815
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1638 GAGTCAGACATCCTGAAGAATTACCTGGCAACCAGAGGTTTGACATGGAAAAACATGTTG 1697

Qy   1816 ACCGAGAGCGCTCGTGGCTCTCCAGCGGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACG 1875
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1698 ACCGAGAGCCCTCGTGGCTCTCCAGCGGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACG 1757

Qy   1876 GGAGACACCGTTCTGTGTTACTGCTGTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAG 1935
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1758 GGAGACACCGTTCTGTGTTACTGCTGTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAT 1817

Qy   1936 TATCGGCAGAGAACATTCCTGCTTCCGAGTTGCCAGTGCCGTAAACATCCCGTCCTGACTGC 1995
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1818 TATCAGCAGAGAACATTCCTGCTTCCGAGTTGCCAGTGCCGTAAACATCCCGTCCTGACTGC 1877

Qy   1996 TACTGGGGCCGTAACTGCCCGCACTCAGGTGAAAGTCAACGCCATGAAATTCAATCAT 2055
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1878 TACTGGGGCCGTAACTGCCCGCACTCAGGTGAAAGTCAACGCCATGAAATTCAATCAT 1937

Qy   2056 ATCTGTGAACAGACAACAAGGTTCAAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCA 2115
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1938 ATCTGTGAACAGACAACAAGGTTCAAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCA 1997

Qy   2116 CTGGAGGTGAAGAGAGCGTGTTTTAAAATACAGAGACAAGCACGTCAAGTGTTTTCAC 2175
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1998 CTGGAGGTGAAGAGAGCGTGTTTTAAAATACAGAGACAAGCACGTCAAGTGTTTTCAC 2057

Qy   2176 AGCCCCCTGAGGGAGGACGGAGGACGAGGGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCTGTG 2235
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2058 AGCCCCCTGAGGGAGGACGGAGGACGAGGGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCTGTG 2117

Qy   2236 GAGCTTTTTACCCTCGAGTGAGATGAGACCCTCCCAGAGCCCCCCAGAGCCCCGCCCCCCTC 2295
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2118 GAGC--TTTACCCTCGAGTGAGATGAGACCCTCCCAGAGCCCCCCAGAGCCCCGCCCCCTC 2175

Qy   2296 CTGGTGAGCGCTGGGCAGGGCGTGTGGTGGCATCAGCAGGACGAGAGACGAAGCCTTTCTGTA 2355
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2176 CTGGTGAGCGCTGGGCAGGGCGTGTGGTGGCATCAGCAGGACGAGAGACGAGAGCGAAGCCTTTCTGTA 2235

Qy   2356 ACATGGCGGCCCGTCCCCGCCGAGAGGGGCAGTTTGCTCTTTTGTACATTTTCCGAAACTAC 2415
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2236 ACATGGCGGCCCGTCCCCGCCGAGAGGGGCAGTTTGCTCTTTTGTACATTTTTCCGAAACTAC 2295

Qy   2416 AGTTAAAGCAGAAGTCTGTTTTCAGGAAAAGTTTCAAGGGAGGGCAAGTTTATCAAA 2475
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2296 AGTTAAAGCCGAAGTCTGTTTTCAGGAAAAGTTTCAAGGGAGGGCAAGTTTATCAAA 2355

Qy   2476 AACATTGTTTCAGGAGAAGGGAGCATTTATCTTTATTTTTATTTTTAATTATAATTACAA 2535
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2356 AACATTGTTTCAGGAGAAGGGAGCATTTATCTTTATCTTTATTTTTAATTATAATTATCCT 2415

Qy   2536 GCTGCTGGGAAAACCAGCATTTAATCTATTTTTTAATTTTTAATTTTGTGTTGTTATC 2595
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2416 GCTGCTGGGAAAACCAGCATTTAATCTATTTTTTAATTTTTAATTTTGTGTTGTTATC 2475

Qy   2596 TTCTAATAGGATTTAAATGTCACAAACTGTAGCACAAATATATAATTATAATTACAA 2655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2476 TTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATATATAATTATAATTACAA 2535

Qy   2656 ATTGACAAAAA 2666
          |||||||||||
Db   2536 ATTGACTAAAA 2546

```
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US026015.
XX
PR   11-SEP-2000; 2000US-00659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   P-PSDB; ABB97233.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis.
XX
PS   Claim 1; SEQ ID NO 57; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth; to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC   and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC   rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a coding sequence of the
CC   invention
XX
SQ   Sequence 3181 BP; 756 A; 830 C; 879 G; 716 T; 0 U; 0 Other;

Query Match      83.8%;  Score 2244.6;  DB 6;  Length 3181;
Best-Local Similarity   92.3%;  Pred. No. 0;
Matches 2447;  Conservative  0;  Mismatches  0;  Indels 125;  Gaps  3;

Qy     16 GGCCGCCAATGTCTCTTGACAGCGGCGGCGGCGGCCGGTTCCGGGTTCGGCGCGGGGGCG     75
          |||||         |||||  |||||||||||||||||||||||||||||||||||||||
Db     21           GGCCGTACCGGTCGGGTCGACGATTCGTTTCCGGGTTCGGCGCGGGGGCG     80

Qy     76 GGGATGTGAATCCGATGGAGCGCCCCGAGGAAGGCAAGCAGTCGCGCCGCCGCAGCCC    135
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GGGATGTGAATCCGATGGAGCGCCCCGAGGAAGGCAAGCAGTCGCGCCGCCGCAGCCC    140

Qy    136 TGGGGACGGCTCCTGCGTCTGGGCGTCGGGCGCGGAGGGCGAGCCGCCGCACGTCTCCTGAGGAAG    195
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 TGGGGACGGCTCCTGCGTCTGGGCGTCGGGCGCGGAGGGCGAGCCGCCGCACGTCCTCCTGAGGAAG    200

Qy    196 CGGGAGTGGACCATCGGGCGCGGAGACGCAGAGGTTGCGACCTTTGCGACCTTTCCTTCCCCAGCAATAAACTG    255
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 CGGGAGTGGACCATCGGGCGCGGAGACGCAGAGGTTGCGACCTTTGCGACCTTTCCTTCCCCAGCAATAAACTG    260

Qy    256 GTCTCTCGGAGATCACTGTAGAATTGTAGTGATGTGAGTGATGAAAATCAGGTTGCAGACACTGGAA    315
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 GTCTCTCGGAGATCACTGTAGAATTGTAGTGATGTGAGTGATGAAAATCAGGTTGCAGACACTGGAA    320

Qy    316 GATACCAGCACCAGTGGAACAGTGATTAACAGCTGAAGGTTGTTAAGAAGCAGACATGC    375
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GATACCAGCACCAGTGGAACAGTGATTAACAGCTGAAGGTTGTTAAGAAGCAGACATGC    380

Qy    376 CCTTTACAGACTGGGGATGTCATCTCACTTGGTACAGGAAGAATGAACCGGAACACAAC    435
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 CCTTTACAGACTGGGGATGTCATCTCACTTGGTACAGGAAGAATGAACCGGAACACAAC    440

Qy    436 GTGGCATACCTCTATGAATCTTTAAGTGAAAAAGCAAGGATACCTCAGGTGCCAGGGCGGA    495
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 GTGGCATACCTCTATGAATCTTTAAGTGAAAAAGCAAGGATACCTCAGGTGCCAGGGCGA    500

Qy    496 GCTAACAAGGAGAAATGTGTTCCATGGGACCCAAAGATACCTCTGGTCTAAAACTCCTGGATCAAGT    555
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 ATGGTGCCTTGCTGTGTGTGTGCCCAGGCTGGTCTAAAACTCCTGGATCAAGT------    551

Qy    556 GGGGCGGATCCCGGGTCCCCGTCGTCGCCGCCGCACTCAGGTGTGCTTGAGGAACCA    615
          ||||
Db    552 --------GATCCTCCACCTTGGCTCCCAAAGTATTGTGATTACA------    590

Qy    616 CAGCCATCAACATCGACGTCAGACCTCTTCCCCACAGCCTCCTGGCCTCTTCCACGGAGCCT    675
Db    591 ----------    590

Qy    676 TCTCCTCGCAGGGCGGAGAGCGGTTCCTCCAGTTGTGGGTCTGGGGTGTGGCATCTCCCT    735
          ||||||||||||||||||||||||
Db    591 ------GGGTCTGGGGTGTGGCATCTCCCT    617

Qy    736 AAAGGAAGTGGTCCCTCTGTGCAAGTGATGAAGTCCAGCTTTGCCTCAGCTCTCCCA    795
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    618 AAAGGAAGTGGTCCCTCTGTGCAAGTGATGAAGTCCAGCTTTGCCTCAGCTCTCCCA    677

Qy    796 GACAGAAAGACTGCGGTCCTTTTCGTCGTTGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCC    855
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    678 GACAGAAAGACTGCGGTCCTTTTCGTCGTTGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCC    737

Qy    856 GTGAAGCAAGAAAATGAGAGGAGGATGGGAACCTTGACCTGAACGGGCAGTTGTTGGTCGCA    915
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    738 GTGAAGAGAGAAAATGAGAGGAGGATGGGAACCTTGACCTGAACGGGCAGTTGTTGGTCGCA    797

Qy    916 CAACCCGGTAGAAATGCCAAACCGTCAGAGGACGTTCAGAGAGCGGCTGGGAAGCCA    975
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 CAACCCGGTAGAAATGCCAAACCGTCAGAGGACGTTCAGAGAGCGGCTGGGAAGCCA    857

Qy    976 GACAAGATGAGGAGGACGCTGACATCATCATCTGCCAGGACCTGCTCGCACGACTGCGTG   1035
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    858 GACAAGATGAGGAGGACGCTGACATCATCATCTGCCAGGACCTGCTCGCACGACTGCGTG    917

Qy   1036 AGTTTGCAGCCCTGCATGCACACGTTCTCGCGGCGTTGCTACTCGGGCTGATGGAGCGC   1095
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    918 AGTTTGCAGCCCTGCATGCACACGTTCTCGCGGCGTTGCTACTCGGGCTGATGGAGCGC    977

Qy   1096 TCGTCCCGTGTCCTACCTGCGCGTGTCCCGTGGAGCGGATCTGTAAAACCACATCCTC   1155
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    978 TCGTCCCGTGTCCTACCTGCGCGTGTCCCGTGGAGCGGATCTGTAAAACCACATCCTC   1037

Qy   1156 AACAACCTCGTGGAAGCATACCTCATCCAGCATCCAGCAACAAGAGTCGCAGTGAACAAGAGAT   1215
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1038 AACAACCTCGTGGAAGCATACCTCATCCAGCATCCAGCAACAAGAGTCGCAGTGAAGAAGAGAT   1097

Qy   1216 GTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGG   1275
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1098 GTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGG   1157

Qy   1276 CGGTCTTTTCTGATGAAGAAGGGAGTTCAGAGGACTGCTGCAGAGCGTGTCAGAGCGTTGAC   1335
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1158 CGGTCTTTTCTGATGAAGAAGGGAGTTCAGAGGACTGCTGCAGAGCGTGTCAGACGTTGAC   1217

Qy   1336 AGTGAGTCCTCAGACATTAGCCAGCATACGTGCGTGTCCCAGAGGCCGAGCCCCACAG   1395
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1218 AGTGAGTCCTCAGACATTAGCCAGCATACGTCGTGTGCCCAGAGGGGCGAGCCCCACAG   1277

Qy   1396 AGGCAGGCGGCGCGCAGCCTCCCACTGCCCAGCACCCGAGCAGTCCGCAGAGAGCCCACAG   1455
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1278 AGGCAGGCGGCGCGCAGCCTCCCACTGCCCAGCACCCGAGCGGCGAGCCCCACAG   1337

Qy   1456 GCCCTGGGGATGCACCCTCCAGTCGTCGTGGACACAGCACAGTCCAGGATTACGTG   1515
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1338 GCCCTGGGGATGCACCCTCCAGTCGTCGTGCACCTGCAGCACAGTCCCAGGATTACGTG   1397

Qy   1516 TGCCCTCTGCAAGGAGCCACGCCGTCGTGCTTCCAGCCATGCCCGACCGG   1575
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1398 TGCCCTCTGCAAGGAGCCACGCCGTCGTGCTTCCAGCGCCATGCCCGACCGG   1457

Qy   1576 AGAGCCGGAGCCGAGCCGCAGCAGGACCCGGCGTGCCCCTCAGCAGTGTCGGGTCGGCGA   1635
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1458 AGAGCCGGAGCCGAGCCGCAGCAGGACCCGGCGTGCCCCTCAGCAGTGTCGGGTCGGCGA------   1517
```

```
Db   1678 CGTGGCTCTCCAGCGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGAGACACCGT 1737
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1887 TCTGTGTTACTGCTGTCTGTGTGGCCTTGCTGGCTGAGCTTCCGTGAGCTGACCTATCAGCAGAA 1946
          |||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1738 TCTGTGTTACTGCTGTGTGTGGCCTGGCAGCTTCCGTGAGCTGACCTATCAGCAGAA 1797
Qy   1947 CATTCCTGCTTCCGAGTTGCCAGTGGCCTAACATCCGTCCTGACTGCTGGGGGCCG 2006
Db   1798 CATTCCTGCTTCCGAGTTGCCAGTGGCCTAACATCCCGTCCTGACTGCTACTGGGGGCCG 1857
Qy   2007 TAACTGCCCACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATATCTGTGAACA 2066
Db   1858 TAACTGCGCACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATATCTGTGAACA 1917
Qy   2067 GACAAGGTTCAAAAACTAAGCATCCAGAGCCCTGAGCAGCTTCAGCACTGGAGGTGAA 2126
Db   1918 GACAAGGTTCAAAAACTAAGCATCCAGAGCCCTGAGCAGCTTCAGCACTGGAGGTGAA 1977
Qy   2127 GAGAGCGTGTTTTTAAAATACAGAGACAAGCACGTGTTTCACAGCCCCCCTGAG 2186
Db   1978 GAGAGCGTGTTTTTAAAATACAGAGACAAGCACGTGTTTCACAGCCCCCCTGAG 2037
Qy   2187 GGAAGGGACGCAGGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCTGTGGAGCTTTTTAC 2246
Db   2038 GGAAGGGACGCAGGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCTGTGGAGC--TTTAC 2095
Qy   2247 CCTCTGAGTGAGACCCCTCCCAGCCCCCGGGGCCCCAGCCCCTCCTGTGAGCGC 2306
Db   2096 CCTCTGAGTGAGACCCCTCCCAGCCCCCGGGGCCCCAGCCCCTCCTGTGAGCGC 2155
Qy   2307 TGGGCAGGGCTCGTGGTGGCATCAGCAGCAGACGAAGCCTTTCTGTAACATGCGGCCG 2366
Db   2156 TGGGCAGGGCTCGTGGTGGCATCAGCAGCAGACGAAGCCTTTCTGTAACATGCGGCCG 2215
Qy   2367 TCCCGCCGAGAGGGCAGTTTGTCTCTTTTGTACATTTTCCGAAACTACAGTTAAAGCAG 2426
Db   2216 TCCTGCCGAGAGGGCAGTTTGTCTCTTTTGTACATTTTCCGAAACTACAGTTAAAGCGG 2275
Qy   2427 AAGTCTGTTTCAGGAAAAGTTTCAAGGAGGAGAGGGCAAGTTTATCAAAAACATTGTTTC 2486
Db   2276 AAGTCTGTTTTCAGGAAAAGTTTCAAGGAGGAGAGGGCAAGTTTATCAAAAAACATTGTTTC 2335
Qy   2487 AGGAGAAGGGAGCATAAGTTACAGCCTACAGACCTACACAATATCCTGCTGCTGGGAA 2546
Db   2336 AGGAGAAGGGAGCATAAGTTACAGCCTACAGACCTACACAATATCCTGCTGCTGGGAA 2395
Qy   2547 AACCACAGCATTTTATCTATTTTTATTTTAATAGGTTTGGTGCTTAATCTTCTAATAAGA 2606
Db   2396 AACCACAGCATTTTATCTATTTTTATTTTAATAGGTTTGGTGCTTAATCTTCTAATAAGA 2455
Qy   2607 TTTAAATGTCACAAACTGAGCACAAATAATAATTACAAATTGACAAAAA 2666
Db   2456 TTTAAATGTCACAAACTGAGCACAAATAATAATATTACAAATTGACTAAAA 2515
```

RESULT 4

ABN59646
ID   ABN59646 standard; cDNA; 3181 BP.
XX
AC   ABN59646;
XX
DT   28-JUN-2002 (first entry)
XX
DE   Novel human coding sequence SEQ ID NO: 57.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.

```
Db    598 TCCCTGTGGCAAGTGATGAAGTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAAGAC 657
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    807 TGCGTCCTTTTCGTCGTTGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAGA 866
Db    658 TGCGTCCTTTTCGTCGTTGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCCGAGAAGAGA 717
Qy    867 AATGAGAGGAGGATGGGACCTTGACCTGAACGGGCAGTTGTTGGTCGCACAACCGCGTAG 926
Db    718 AATGAGAGGAGGATGGGGACCTTGACCTGAACGGGCAGTTGTTGGTCGCACAACCGCGTAG 777
Qy    927 AAATGCCCAAACCGTCCACGAGACGCGGCTGGGAAGCCAGACAAGATGA 986
Db    778 AAATGCCCAAACCGTCCACGAGACGCGGCTGGGAAGCCAGACAACAAGATGGA 837
Qy    987 GGGAAGACGCTGACATGCATCATCTGCCAGGACCTGCTCGCACGAGCTGCGTGAGTTTGCAGCC 1046
Db    838 GGGAAGACGCTGACATGCATCATCTGCCAGGACCTGCTGCTCGCACGAGCTGCGTGAGTTTGCAGCC 897
Qy   1047 CTGCATGCACACGTTCTGCGGCTTCGACTGAACGGGCTGCTACTCGGCTCGTCGTCCCTGTG 1106
Db    898 CTGCATGCACACGTTCTGCGGCTTCGACTGAACGGGCTGCTACTCGGCTCGTCGTCCCTGTG 957
Qy   1107 TCCTTACCTGCGCCGTGTCCCGTCGTGGGCGGATCGTAAAAACCACATCCTCAACAACCTCGT 1166
Db    958 TCCTTACCTGCGCCGTGTCCCGTCGTGGGCGGATCGTAAAAACCACATCCTCAACAACCTCGT 1017
Qy   1167 GGAAGCATACCTCATCCAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAGTAT 1226
Db   1018 GGAAGCATACCTCATCCAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAGTAT 1077
Qy   1227 GGATGCCAGGAATAAAATCACTCAAGACATGCTCAGCGCCAAAGTCAGGCGGGTCTTTTTC 1286
Db   1078 GGATGCCAGGAATAAAATCACTCAAGACATGCTCAGCGCCAAAGTCAGGCGGGTCTTTTTC 1137
Qy   1287 TGATGAAGAAGGGAGTTCAGAGGACCTGCTGCAGGAGTGTCGAGACGTTGACGTCCTC 1346
Db   1138 TGATGAAGAAGGGAGTTCAGAGGACCTGCTGCAGGAGTGTCGAGACGTTGACGTGAGTCCTC 1197
Qy   1347 AGACATTAGCCAGCCATACGTCGTGTCCGCAGTGTCCTGAGTACAGAGGCAGGGCGGC 1406
Db   1198 AGACATTAGCCAGCCATACGTCGTGTCCGCAGTGTCCTGAGTACAGAGGCAGGGCGGC 1257
Qy   1407 GCAGCCTCCCCACTGCCCAGACCCCAGAGGGCGAGCCAGGACCCCCACAGCCCCTGGGGGA 1466
Db   1258 GCAGCCTCCCCACTGCCCAGACCCCAGAGGGCGAGCCAGGACCCCCACAGCCCCTGGGGGA 1317
Qy   1467 TGCACCCCTCCAGTGCCAGTCTCAGCCTCAGGATTACGTGTGCCCCTCTGCA 1526
Db   1318 TGCACCCCCTCCAGTGCCAGTCTCAGCCTCAGGATTACGTGTGCCCCTCTGCA 1377
Qy   1527 AGGAAGCCACGCCCGTGCACCCCGTCACCCGGCAGCCGGGAGCGGAGCG 1586
Db   1378 AGGAAGCCACGCCCGTGCACCCCGTCACCCGGCAGCCGGGAGCGGAGCG 1437
Qy   1587 CGAGCAGGAGGACCCCGTCCGCCCTGTCGCGGTTCGCGGGTTCCTGCCA 1646
Db   1438 CGAGCAGGAGGACCCCGTCCGCCCTGTCGCGGTTCGCGGTTCTGCCA 1497
Qy   1647 CCTGTACTGGGGCTGCACCCGGACCCGGCTGCCTGGCCCGGTTTTGTGAGCT 1706
Db   1498 CCTGTACTGGGGCTGCACCCGGACCCGGCTGCCTGGCCCGGTTTTGTGAGCT 1557
Qy   1707 CAACCTGGGTGACAGTGTCTGACACGGCGTGCTGAACAACAACAGCTACGAGTCAGACAT 1766
Db   1558 CAACCTGGGTGACAGTGTCTGACACGGCGTGCTGAACAACAACAGCTACGAGTCAGACAT 1617
Qy   1767 CCTGAAGAATTACCTGCAACCAGAGGTTTGACATGGAAAAACATGTTGACCGGAGACCT 1826
Db   1618 CCTGAAGAATTACCTGCAACCAGAGGTTTGACATGGAAAAACATGTTGACCGGAGACCT 1677
Qy   1827 CGTGGCTCTCCAGCGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGAGACACCGT 1886
```

CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 3138 BP; 738 A; 824 C; 873 G; 703 T; 0 U; 0 Other;

Query Match   84.4%; Score 2262.4; DB 4; Length 3138;
Best Local Similarity   93.0%; Pred. No. 0;
Matches 2454; Conservative   0; Mismatches   61; Indels 125; Gaps   3;

Qy   27 CTCTTGACAGCGCGCGGCGGCGCGCCGAGGAAGGCAGAGCAGTCGCCGCGCCGGTTCGCGCGGCGGGCGGGGGATGTGAAT   86
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CTCTTGACAGCGCGCGGCGGCGCGCCGAGGAAGGCAGAGCAGTCGCCGCGCCGGTTCGCGCGGCGGGCGGGGGATGTGAAT   60

Qy   87 CCCGATGGAGCGGCCCGAGGAAGGCAGAGCAGTCGCCGCGCCGCCGCCGGTTCCTCCTGAGGAAGGGGAGTGGAC 146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CCCGATGGAGCGGCCCGAGGAAGGCAGAGCAGTCGCCGCGCCGCCGCCGGTTCCTCCTGAGGAAGGGGAGTGGAC 120

Qy   147 CCTGCGTCTGGGCGGGAGGAGGGGGAGCCGCCGAGTCGCCGGTTCCTCCTGAGCAATAAACTGGTCTCTGGAGA 206
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCTGCGTCTGGGCGGGAGGAGGGGGAGCCGCCGAGTCGCCGGTTCCTCCTGAGCAATAAACTGGTCTCTGGAGA 180

Qy   207 CATCGGGCGGAGAGGAGGGTTGCGACCTTTCCTTCCCAGCAATAAACTGGTCTCTGGAGA 266
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CATCGGGCGGAGAGGAGGGTTGCGACCTTTCCTTCCCAGCAATAAACTGGTCTCTGGAGA 240

Qy   267 TCACTGTAGAATTGTGAGTGAGTGGATGAAAATCAGGTCAGGTGACACTGGAGATACCAGCAC 326
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TCACTGTAGAATTGTGAGTGAGTGGATGAAAATCAGGTCAGGTGACACTGGAGATACCAGCAC 300

Qy   327 CAGTGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGACATGCCCTTTACAGAC 386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CAGTGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGACATGCCCTTTACAGAC 360

Qy   387 TGGGGATGTCATCATCTACTTGGTGTACAGGAAGAATGAACCGGAACACAACGTGGCATACCT 446
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGGGGATGTCATCATCTACTTGGTGTACAGGAAGAATGAACCGGAACACAACGTGGCATACCT 420

Qy   447 CTATGAATCTTCTTAAGTGAAAAGCAAGGCATGAAGCATGAACACAAGAATCCTTTGAGCTAACAAGGA 506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CTATGAATCTTCTTAAGTGAAAAGCAAGGCATGAAGCATGAACACAAGAATCCTTTGAGCTAACAAGGA 480

Qy   507 AAATGTGTTCCATGGGACCAAAGATACCTCAGGTGCAGGTGAGCTGAGGGGCCGATCC 566
          |||||||||||||||||||||||||||||||||||||||| ||||
Db   481 CTGTGTTGCCCAGCTGGTGGTCTAAAACTCCTGGGATCAAGT------------------------- 520

Qy   567 CCGGGTCCCTCCGTCGTCGCCGCCACTCCAGGTGTGCTTTGAGGAACCACCAGCCATCAAC 626
          |||||||||||||||||||||||| ||| ||
Db   521 CC---GATCCTCCACCTTGGCTCCCCAAAGTATTGTGATTACA----------------- 559

Qy   627 ATCGACGTCAGACCTCTTCCCCACGAGCCTCGGCCCTTCTCCACGGAGCCTTCTCCTGCAGG 686

Db   560 --- 559

Qy   687 GCCGAGCGGTTCCTCCAGTTGTGGGTCTGGGGTGGTCGCATCTCCCCTAAAGGAAGTGG 746
          ||| |||||| ||||| |||| |||
Db   560 ----------------GGTCTGGGGGTGGTGGCATCTCCCCTAAAGGAAGTGG 597

Qy   747 TCCCTCTGTGCAAGTGACGTGATGAAGTCCAGCTTTGCTCCAGTCCTCCCCAGACAGAAAGAC 806

Db   2249 TGGTGAGCGGCTGGGCAGGGGCTGCTGGTGCCAGTGCCAGTTTGCTCTTTTGTACATTTCCGAAACTACA 2308
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2357 CATGGCGGCCGTCCCGCCGAGAGGGGCAGTTTGCTCTTTTGTACATTTCCGAAACTACA 2416
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2309 CATGGCGGCCGTCCCGCCGAGAGGGGCAGTTTGCTCTTTTGTACATTTCCGAAACTACA 2368

Qy   2417 GTTAAAGCAGAAGTCTGTTTCAGGAAAAGTTCAAGGGAGAAGGGCAAGTTATCAAAA 2476
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2369 GTTAAAGCAGAAGTCTGTTTCAGGAAAAGTTCAAGGGAGAAGGGCAAGTTATCAAAA 2428

Qy   2477 ACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTACAGGAGCTACACAATATCCTG 2536
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2429 ACATTGTTTCAGGAGAAGGGGAGCATAAGTTTACAGCCTACAGGAGCTACACAATATCCTG 2488

Qy   2537 CTGCTGGGAAAACCACCAGCATTTATCTCTATTTTTTTATTTTTTAATAGGTTTGTGCTTATCT 2596
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2489 CTGCTGGGAAAACCACCAGCATTTATCTCTATTTTTTTATTTTTTAATAGGTTTGTGCTTATCT 2548

Qy   2597 TCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATAATAATTTATAATTTACAAA 2656
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2549 TCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATAATAATTTATAATTTACAAA 2608

Qy   2657 TTGACAAAAA 2666
          ||| ||||||
Db   2609 TTGACTAAAA 2618

RESULT 3
AAH14556
ID   AAH14556 standard; cDNA; 3138 BP.
XX
AC   AAH14556;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:12127.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999; 99JP-00248036.
PR   27-AUG-1999; 99JP-00300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 12127; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the

XX
AC   AAH14542;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:12099.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000BP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000;   2000JP-00118776.
PR   02-MAY-2000;   2000JP-00183767.
PR   09-JUN-2000;   2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 12099; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 2639 BP; 620 A; 706 C; 760 G; 553 T; 0 U; 0 Other;

Query Match   95.8%;   Score 2566.4;   DB 4;   Length 2639;
Best Local Similarity   98.6%;   Pred. No. 0;
Matches 2613;   Conservative   0;   Mismatches   1;   Indels   36;   Gaps   1;

Qy   17 GCCGGCAATGTCTCTTGACAGCGGCGGCCGCGGCCGCGGCCGCAGCCGCAGCCGGTTCGGGCGGTTCGGGCGGCGGCCAGCCGCAGCCGGTTCGGGCGGGGCGG 76
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5 GCCGGCAATGTCTCTTGACAGCGGCGGCCGCGGCCGCGGCCGCAGCCGCAGCCGGTTCGGGCGGTTCGGGCGGCGGCCAGCCGCAGCCGGTTCGGGCGGGGCGG 64

Qy   77 GGATGTGAATCCCATGGAGCGGCCGGCCGGAGGAAGGCAAGCAGTCGCCGGCCGCCGCCAGCCCT 136
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 GGATGTGAATCCCATGGAGCGGCCGGCCGGAGGAAGGCAAGCAGTCGCCGGCCGCCGCCAGCCCT 124

1681 GGCTGCCTGGCCCCGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTGACGGCGTGCTG 1740
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1681 GGCTGCCTGGCCCCGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTGACGGCGTGCTG 1740

1741 AACAACAACAGCTACGAGTCAGACATCCTGAAGAATTACCTGGCAACCAGAGGTTTGACA 1800
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1741 AACAACAACAGCTACGAGTCAGACATCCTGAAGAATTACCTGGCAACCAGAGGTTTGACA 1800

1801 TGGAAAAACATGTTGACCGGAGAGCCTCGTGGCTCCCAGCGGGGAGTGTTTCTGCTGTCT 1860
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1801 TGGAAAAACATGTTGACCGGAGAGCCTCGTGGCTCCCAGCGGGGAGTGTTTCTGCTGTCT 1860

1861 GATTACAGAGCTCACGGAGACACCCGTTCTGTTACTGCTGCTGTGCCCTGCGCAGCTTCCGT 1920
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1861 GATTACAGAGCTCACGGAGACACCCGTTCTGTTACTGCTGCTGTGCCCTGCGCAGCTTCCGT 1920

1921 GAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACA 1980
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1921 GAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACA 1980

1981 TCCCGTCCTGACTGCTACTGGGGCCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCC 2040
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1981 TCCCGTCCTGACTGCTACTGGGGCCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCC 2040

2041 ATGAAATTCAATCATATCATCTGTGAACAGACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCT 2100
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2041 ATGAAATTCAATCATATCATCTGTGAACAGACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCT 2100

2101 GAGCAGCTTTCAGCACTGGAGTGAAGAGAGAGTGTTGTTTTAAAATACAGAGACAAGCACG 2160
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2101 GAGCAGCTTTCAGCACTGGAGTGAAGAGAGAGCGTGTTTTTAAAATACAGAGACAAGCACG 2160

2161 TCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGGACGCAGGGTCTCCGACAGGGTGCTCTGG 2220
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2161 TCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGGACGCAGGGTCTCCGACAGGGTGCTCTGG 2220

2221 GGTGACTCTTCTGTGGAGCTTTTACCCTCTGAGTGAGACCCTCCCAGAGCCCCGGGGG 2280
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2221 GGTGACTCTTCTGTGTGGAGCTTTTACCCTCTGAGTGAGACCCTCCCAGAGCCCCGGGGG 2280

2281 CCGCAGCCGCCGCCCCTCCTGGAGCGCGTGGGCAGGGCTCGTGGTGGCATCAGCAGCAGAGA 2340
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2281 CCGCAGCCGCCGCCCCTCCTGGTGAGCGCGTGGGCAGGGCTCGTGGTGGCATCAGCAGCAGAGA 2340

2341 CGAAGCCTTTCTGTAACATGCGGCCGCCTCCCGCCGAGAGGGGCAGTTTGCTCTTTGTAC 2400
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2341 CGAAGCCTTTCTGTAACATGCGGCCGCCTCCCGCCGAGAGGGGCAGTTTGCTCTTTTGTAC 2400

2401 ATTTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTTCAGGGAAAAGTTTCAAGGGAGAAG 2460
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2401 ATTTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTTCAGGGAAAAGTTTCAAGGGAGAAG 2460

2461 GGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGACATAAGTTTACAGCCTACAGGA 2520
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2461 GGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTACAGGA 2520

2521 CGTACCAATATCCTGCTGCTGGGAAAACCACACAGCATTTATCTATTTTTATTTTAATA 2580
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2521 CGTACCAATATCCTGCTGCTGGGAAAACCACACAGCATTTATCTATTTTTATTTTAATA 2580

2581 GGTTTGGTTGCTTAATCTTCTAATAAGATTTAAAATGTCACAAACTGTAGCACAAATAATATA 2640
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2581 GGTTTGGTTGCTTAATCTTCTAATAAGATTTAAAATGTCACAAACTGTAGCACAAATAATATA 2640

2641 ATTTATAATTTACAAATTGACAAAAAAAAAAAAAAAAAA 2679
     |||||||||||||||||||||||||||||||||||||||
2641 ATTTATAATTTACAAATTGACAAAAAAAAAAAAAAAAAA 2679

RESULT 2
AAH14542
ID   AAH14542 standard; cDNA; 2639 BP.

```

```
CC   checkpoint protein Chfr (see AAB20219) having a forkhead associated
CC   domain (FHA) and a ring finger domain. The protein is required for
CC   regulation of the transition of cells from prophase to metaphase during
CC   mitosis. It has ubiquitin-protein ligase activity. The Chfr checkpoint
CC   was evident in primary human cells, but was inactivated in 4 of 8 human
CC   cancer cell lines. In U2OS cells, a mutation was identified, which caused
CC   a Val to Met amino acid substitution in the highly conserved C-terminal
CC   Cys-rich region of the Chfr protein. In the absence of the Chfr
CC   checkpoint, cells subjected to mitotic stress condensed their chromosomes
CC   despite failing to separate their chromosomes. Chfr may monitor
CC   centrosome separation. Inactivation of the chfr gene in human cancer is
CC   theorized to underlie the increased sensitivity of cancer cells to
CC   antimitotic drugs. Nucleic acids comprising the present sequence, or
CC   sequences encoding at least amino acids 31-103, 303-346, 476-641 (or
CC   their antisense sequences) are claimed. The chfr cDNA was isolated from
CC   an expressed sequence tag database sequence for cDNAs with FHA motifs.
CC   Claimed methods of determining the tumourigenic potential of a cell
CC   comprise examining the cell for the presence of chfr nucleic acid
CC   (absence or mutation indicating predisposition to tumourigenesis upon
CC   exposure to mitotic stress). Also claimed is a diagnostic agent, e.g. an
CC   antisense fragment of the present sequence, that binds to the chfr
CC   nucleic acid, and a diagnostic kit for detecting tumourigenic potential
CC   of a cell. A composition which inhibits the biological activity of Chfr
CC   may comprise a ligand selected from an antibody or its fragment. The Chfr
CC   inhibitor is used in a claimed method of retarding the growth of a cancer
CC   cells. Chfr polypeptides are useful in screening for drugs which can
CC   inhibit the activity of Chfr in a cancer cell, rendering the cell more
CC   sensitive to additional antitumour therapies
XX
SQ   Sequence 2679 BP; 650 A; 715 C; 764 G; 550 T; 0 U; 0 Other;

  Query Match   100.0%; Score 2679; DB 5; Length 2679;
  Best Local Similarity   100.0%; Pred. No. 0;
  Matches 2679; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 AAGAATTCGGCACGAGGCCGCAATGTCTCTTGACAGCGGCGGCGGCGCAGCCGGTTCCGG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AAGAATTCGGCACGAGGCCGCAATGTCTCTTGACAGCGGCGGCGGCGCAGCCGGTTCCGG 60

Qy   61 GTTCGGCGCGGGGCGGGGATGTGAATCCCGATGGAGCGGCCCGAGGAAGGCAAGCAGTCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GTTCGGCGCGGGGCGGGGATGTGAATCCCGATGGAGCGGCCCGAGGAAGGCAAGCAGTCG 120

Qy   121 CCGCCGCCGCAGCCCTGGGGACGGCTCCTGCGTCTGGGCGCGGAGGAGGGCGAGCCGCAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCGCCGCCGCAGCCCTGGGGACGGCTCCTGCGTCTGGGCGCGGAGGAGGGCGAGCCGCAC 180

Qy   181 GTCCTCCTGAGGAAGCGGGAGTGGACCATCGGGCGGAGACGAGGTTGCGACCTTTCCTTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTCCTCCTGAGGAAGCGGGAGTGGACCATCGGGCGGAGACGAGGTTGCGACCTTTCCTTC 240

Qy   241 CCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGT 300

Qy   301 CAGGTGACACTGGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CAGGTGACACTGGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTT 360

Qy   361 AAGAAGCAGACATGCCCTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGAAGCAGACATGCCCTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAAT 420

Qy   421 GAACCGGAACACAACGTGGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGGCATGACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAACCGGAACACAACGTGGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGGCATGACA 480

Qy   481 CAAGAATCCTTTGAAGCTAACAAGGAAAATGTGTTCCATGGGACCAAAGATACCTCAGGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CAAGAATCCTTTGAAGCTAACAAGGAAAATGTGTTCCATGGGACCAAAGATACCTCAGGT 540

Qy   541 GCAGGTGCAGGGCGAGGGGCCGATCCCCGGGTCCCTCCGTCGTCGCCCGCCACTCAGGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GCAGGTGCAGGGCGAGGGGCCGATCCCCGGGTCCCTCCGTCGTCGCCCGCCACTCAGGTG 600

Qy   601 TGCTTTGAGGAACCACAGCCATCAACATCGACGTCAGACCTCTTCCCCACAGCCTCGGCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGCTTTGAGGAACCACAGCCATCAACATCGACGTCAGACCTCTTCCCCACAGCCTCGGCC 660

Qy   661 TCTTCCACGGAGCCTTCTCCTGCAGGGCGAGAGCGTTCCTCCAGTTGTGGGTCTGGGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TCTTCCACGGAGCCTTCTCCTGCAGGGCGAGAGCGTTCCTCCAGTTGTGGGTCTGGGGGT 720

Qy   721 GGTGGCATCTCCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GGTGGCATCTCCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTT 780

Qy   781 GCCTCAGCTCTCCCAGACAGAAAGACTGCGTCCTTTTCGTCGTTGGAACCCCAGGATCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCCTCAGCTCTCCCAGACAGAAAGACTGCGTCCTTTTCGTCGTTGGAACCCCAGGATCAG 840

Qy   841 GAGGATTTGGAGCCCGTGAAGAAGAAAATGAGAGGAGATGGGGACCTTGACCTGAACGGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 GAGGATTTGGAGCCCGTGAAGAAGAAAATGAGAGGAGATGGGGACCTTGACCTGAACGGG 900

Qy   901 CAGTTGTTGGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAGGACGTCAGAGCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 CAGTTGTTGGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAGGACGTCAGAGCA 960

Qy   961 GCGGCTGGGAAGCCAGACAAGATGGAGGAGACGCTGACATGCATCATCTGCCAGGACCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GCGGCTGGGAAGCCAGACAAGATGGAGGAGACGCTGACATGCATCATCTGCCAGGACCTG 1020

Qy   1021 CTGCACGACTGCGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCGGCTTGCTACTCG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CTGCACGACTGCGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCGGCTTGCTACTCG 1080

Qy   1081 GGCTGGATGGAGCGCTCGTCCCTGTGTCCTACCTGCCGCTGTCCCGTGGAGCGGATCTGT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GGCTGGATGGAGCGCTCGTCCCTGTGTCCTACCTGCCGCTGTCCCGTGGAGCGGATCTGT 1140

Qy   1141 AAAAACCACATCCTCAACAACCTCGTGGAAGCATACCTCATCCAGCATCCAGACAAGAGT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AAAAACCACATCCTCAACAACCTCGTGGAAGCATACCTCATCCAGCATCCAGACAAGAGT 1200

Qy   1201 CGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTG 1260

Qy   1261 CAGCCCAAAGTCAGGCGGTCTTTTTCTGATGAAGAAGGGAGTTCAGAGGACCTGCTGGAG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CAGCCCAAAGTCAGGCGGTCTTTTTCTGATGAAGAAGGGAGTTCAGAGGACCTGCTGGAG 1320

Qy   1321 CTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAG 1380

Qy   1381 TGTCCTGAGTACAGAAGGCAGGCGGCGCAGCCTCCCCACTGCCCAGCACCCGAGGGCGAG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TGTCCTGAGTACAGAAGGCAGGCGGCGCAGCCTCCCCACTGCCCAGCACCCGAGGGCGAG 1440

Qy   1441 CCAGGAGCCCCACAGGCCCTGGGGGATGCACCCTCCACGTCCGTCAGCCTGACGACAGCA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CCAGGAGCCCCACAGGCCCTGGGGGATGCACCCTCCACGTCCGTCAGCCTGACGACAGCA 1500

Qy   1501 GTCCAGGATTACGTGTGCCCTCTGCAAGGAAGCCACGCCCTGTGCACCTGCTGCTTCCAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GTCCAGGATTACGTGTGCCCTCTGCAAGGAAGCCACGCCCTGTGCACCTGCTGCTTCCAG 1560

Qy   1561 CCCATGCCCGACCGGAGAGCGGAGCGCGAGCAGGACCCGCGTGTCGCCCCTCAGCAGTGT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CCCATGCCCGACCGGAGAGCGGAGCGCGAGCAGGACCCGCGTGTCGCCCCTCAGCAGTGT 1620

Qy   1621 GCGGTCTGCCTGCAGCCTTTCTGCCACCTGTACTGGGGCTGCACCCGGACCGGCTGCTAC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GCGGTCTGCCTGCAGCCTTTCTGCCACCTGTACTGGGGCTGCACCCGGACCGGCTGCTAC 1680
```

                GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 14, 2004, 11:05:18 ; Search time 1484 Seconds
                                     (without alignments)
                          7669.085 Million cell updates/sec

Title: US-10-048-046-1
Perfect score: 2679
Sequence: 1 aagaattcggcacgaggccg.........acaaaaaaaaaaaaaaaaaa 2679

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_29jan04:*
            1: geneseqn1990s:*
            2: geneseqn1990s:*
            3: geneseqn2000s:*
            4: geneseqn2001as:*
            5: geneseqn2001bs:*
            6: geneseqn2002s:*
            7: geneseqn2003as:*
            8: geneseqn2003bs:*
            9: geneseqn2003cs:*
           10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

### SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 2679 | 100.0 | 2679 | 5 | AAF30352 | Aaf30352 Human chf |
| 2 | 2566.4 | 95.8 | 2639 | 4 | AAH14542 | Aah14542 Human cDN |
| 3 | 2262.4 | 84.4 | 3138 | 4 | AAH14556 | Aah14556 Human cDN |
| 4 | 2244.6 | 83.8 | 3181 | 6 | ABN59646 | Abn59646 Novel hum |
| 5 | 1995 | 74.5 | 1995 | 5 | AAF89709 | Aaf89709 Nucleotid |
| c 6 | 1294.2 | 48.3 | 2259 | 4 | AAI88903 | Aai88903 Human pol |
| 7 | 945 | 35.3 | 1311 | 4 | AAS25843 | Aas25843 Human cDN |
| 8 | 945 | 35.3 | 1311 | 7 | ABX73184 | Abx73184 Human nov |
| 9 | 876.2 | 32.7 | 2186 | 7 | ADA52592 | Ada52592 Human cod |
| 10 | 727 | 27.1 | 816 | 4 | AAH06828 | Aah06828 Human cDN |
| 11 | 583 | 21.8 | 824 | 4 | AAH06846 | Aah06846 Human cDN |
| 12 | 536.6 | 20.0 | 693 | 4 | AAS26304 | Aas26304 Human cDN |
| 13 | 536.6 | 20.0 | 693 | 7 | ABX73645 | Abx73645 Human nov |
| 14 | 495.8 | 18.5 | 575 | 3 | AAA44336 | Aaa44336 Human sec |
| c 15 | 488.8 | 18.2 | 518 | 4 | AAH11859 | Aah11859 Human cDN |
| 16 | 450 | 16.8 | 825 | 4 | AAK58374 | Aak58374 Human imm |
| 17 | 394.4 | 14.7 | 476 | 8 | ACH14968 | Ach14968 Human adu |
| 18 | 310.2 | 11.6 | 357 | 4 | AAS26143 | Aas26143 Human cDN |
| 19 | 310.2 | 11.6 | 357 | 7 | ABX73484 | Abx73484 Human nov |
| 20 | 273.6 | 10.2 | 449 | 8 | ACH26450 | Ach26450 Human adu |
| 21 | 256.4 | 9.6 | 354 | 4 | AAS26563 | Aas26563 Human cDN |
| 22 | 256.4 | 9.6 | 354 | 4 | AAS26144 | Aas26144 Human cDN |
| 23 | 256.4 | 9.6 | 354 | 7 | ABX73904 | Abx73904 Human nov |
| 24 | 256.4 | 9.6 | 354 | 7 | ABX73485 | Abx73485 Human nov |
| 25 | 170 | 6.3 | 13836 | 5 | ABA16633 | Aba16633 Human ner |
| c 26 | 82.2 | 3.1 | 617 | 6 | ABN60913 | Abn60913 Human can |
| 27 | 60 | 2.2 | 60 | 6 | ABN33472 | Abn33472 Human spl |
| 28 | 52.4 | 2.0 | 2000 | 7 | ADA71938 | Ada71938 Rice gene |
| c 29 | 51.6 | 1.9 | 2000 | 7 | ADA71938 | Ada71938 Rice gene |
| 30 | 49.4 | 1.8 | 612 | 4 | AAH71471 | Aah71471 Human cer |
| 31 | 48.2 | 1.8 | 434 | 7 | ABX38360 | Abx38360 Bovine ES |
| c 32 | 47.8 | 1.8 | 7455 | 6 | ABL33759 | Abl33759 Human imm |
| c 33 | 47.8 | 1.8 | 309 | 5 | ABV44994 | Abv44994 Human pro |
| 34 | 47.8 | 1.8 | 312 | 7 | ABX43249 | Abx43249 Bovine ES |
| 35 | 47.8 | 1.8 | 474 | 5 | ABV56492 | Abv56492 Human pro |
| 36 | 47.8 | 1.8 | 556 | 5 | ABV40163 | Abv40163 Human pro |
| c 37 | 47.8 | 1.8 | 556 | 5 | ABV40063 | Abv40063 Human pro |
| 38 | 47.8 | 1.8 | 556 | 5 | ABV42105 | Abv42105 Human pro |
| c 39 | 47.8 | 1.8 | 556 | 5 | ABV43601 | Abv43601 Bovine ES |
| c 40 | 47.2 | 1.8 | 424 | 4 | ABX46053 | Abx46053 Bovine ES |
| 41 | 47.2 | 1.8 | 612 | 4 | AAH71471 | Aah71471 Human cer |
| 42 | 47.2 | 1.8 | 2122 | 7 | ACC79105 | Acc79105 Human sec |
| c 43 | 47.2 | 1.8 | 6215 | 6 | ABL33190 | Abl33190 Human imm |
| 44 | 47 | 1.8 | 2497 | 6 | AAC61197 | Aac61197 Human cDN |
| 45 | 46.6 | 1.7 | 1290 | 8 | AAD57083 | Aad57083 Human ald |

### ALIGNMENTS

```
RESULT 1
AAF30352
AAF30352
ID   AAF30352 standard; cDNA; 2679 BP.
XX
AC   AAF30352;
XX
DT   14-MAY-2001  (first entry)
XX
DE   Human chfr cDNA encoding checkpoint with FHA and ring finger protein.
XX
KW   Checkpoint with forkhead associated domain and ring finger; Chfr; human;
KW   mitosis; cell cycle; tumour; diagnosis; antitumour; drug screening;
KW   ubiquitin-protein ligase; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             91..2085
FT                   /*tag= a
FT                   replace(1828,A)
FT   mutation        /*tag= b
FT                   /note= "alters codon GTG (Val) to ATG (Met) in chfr from
FT                   human cancer cell line U2OS"
XX
PN   WO200109150-A2.
XX
PD   08-FEB-2001.
XX
PF   14-JUN-2000; 2000WO-US016391.
XX
PR   29-JUL-1999; 99US-0146194P.
XX
PA   (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI   Halazonetis T, Scolnick D;
XX
DR   WPI; 2001-182927/18.
DR   P-PSDB; AAB20219.
XX
PT   Novel nucleic acid sequence of mitotic checkpoint gene encoding a
PT   checkpoint with forkhead-associated domain and ring finger protein, for
PT   diagnosing tumorigenic cells and in screening for anticancer drugs.
XX
PS   Claim 2(a); Fig 4A-D; 85pp; English.
XX
CC   The present sequence is that of cDNA encoding the human mitotic
```

```
Qy   2122  GTGAAGAGAGCGTCGTTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTCACAGCCCC  2181
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089  GTGAAGAGAGCGTGTTTTTAAAATACAGAGACACGTCAAGGTGTTTCACAGCCCC  1148

Qy   2182  CTGAGGGAAGGGACGCAGGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCTGTGGAGCTT  2241
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149  CTGAGGGAAGGGACGCAGGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCTGTGGAGCTT  1208

Qy   2242  TTTACCCTCTGAGTGAGACCCTCCCAGAGCCCCGGGGGCCGCAGCCCGCCCCTCCTGGTG  2301
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209  TTTACCCTCTGAGTGAGACCCTCCCAGAGCCCCGGGGGCCGCAGCCCGCCCCTCCTGGTG  1268

Qy   2302  AGCGCTGGGCAGGCTCGTGGTGGCATCAGCAGCAGAGACGAAGCCTTTCTGTAAACATGC  2361
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269  AGCGCTGGGCAGGGCTCATGGTGGCATCAGCAGCAGAGACGAAGCCTTTCTGTAACATGC  1328

Qy   2362  GGCCGTCCCGCCGAGAGGGGCAGTTTGCTCTCTTTGTACATTTCCGAAACTACAGTTAA  2421
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329  GGCCGTCCCGCCGAGAGGGGCAGTTTGCTCTCTTTGTACATTTCCGAAACTACAGTTAA  1388

Qy   2422  AGCAGAAGTCTGTTTCAGGAAAAGTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATT  2481
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389  AGCAGAAGTCTGTTTCAGGAAAAGTTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATT  1448

Qy   2482  GTTTCAGGAGAAGGGAGCATAAGTTACAGCTACAGACGTACACAATATCCTGCTGCT  2541
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1449  GTTTCAGGAGAAGGGAGCATAAGTTTACAGCTACAGACGTACACAATATCCTGCTGCT  1508

Qy   2542  GGGAAAACCACCAGACATTTTATCTATTTTTTATTTTAATAGGTTGGTGCTTATCTTCTAA  2601
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1509  GGGAAAACCACCAGACATTTTATCTATTTTTTATTTTTAATAGGTTGGTGCTTATCTTCTAA  1568

Qy   2602  TAAGATTTAAAATGTCACAAACTGTAGCACAAATAATATATAATTTATAAATTGAC  2661
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1569  TAAGATTTAAATGTCACAAACTGTAGCACAAATATATATAATTTATAATTTACAAATTGAC  1628

Qy   2662  AAAAA  2666
            |||||
Db   1629  TAAAA  1633
```

Search completed: May 14, 2004, 17:19:06
Job time : 16450 secs

SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
                Kodaira,H., Furuya,T., Takahashi,H., Kikkawa,E., Omura,Y., Abe,K.,
                Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
                Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
                Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
                Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
                Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
                Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
                Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE   NEDO human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2257)
AUTHORS   Isogai,T. and Yamamoto,J.
TITLE   Direct Submission
JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
                Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
                (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
                Economy, Trade and Industry of Japan; cDNA full insert sequencing:
                Research Association for Biotechnology (RAB); cDNA library
                construction: Helix Research Institute (HRI) (supported by Japan
                Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
                HRI, and Biotechnology Center, National Institute of Technology and
                Evaluation; clone selection for full insert sequencing: HRI and
                RAB; annotation: HRI and RAB.
FEATURES   Location/Qualifiers
     source   1..2257
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="BRAMY2022454"
                     /tissue_type="amygdala"
                     /clone_lib="BRAMY2"
                     /note="cloning vector: pME18SFL3"
ORIGIN

Query Match   35.9%; Score 961.8; DB 9; Length 2257;
Best Local Similarity   99.8%; Pred. No. 4.1e-210;
Matches 963; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1702 GAGCTCAACCTGGGTGACAAGTGTCTGGACGGCGTGCTGAACAACAGCTACGAGTCA 1761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   669 GAGCTCAACCTGGGTGACAAGTGTCTGGACGGCGTGCTGAACAACAGCTACGAGTCA 728

Qy   1762 GACATCCTGAAGAATTACCTGGCAACCAGCAGAGTTTGACATGTGGAAAAACATGTTGACCGAG 1821
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   729 GACATCCTGAAGAATTACCTGGCAACCAGCAGAGTTTGACATGTGGAAAAACATGTTGACCGAG 788

Qy   1822 AGCCTCGTGGCCTCCAGCGGGGGAGTGTTTCTGCGTCTGATTACAGAGTCACGGGAGAC 1881
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   789 AGCCTCGTGGCCTCCAGCGGGGGAGTGTTTCTGCGTCTGATTACAGAGTCACGGGAGAC 848

Qy   1882 ACCGGTTCTGTTACTGCCTGGCCTGGCCGCAGCTTCCGTGACGCTTCCGTAACATCCGG 1941
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   849 ACCGGTTCTGTTACTGCCTGGCCTGGCCGCAGCTTCCGTGACGCTTCCGTAACATCCGG 908

Qy   1942 CAGAACAATTCCTGCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCTGACTACTGG 2001
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   909 CAGAACAATTCCTGCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCTGACTACTGG 968

Qy   2002 GGCCGTAACTGCCGACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATATCTGT 2061
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   969 GGCCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATATCTGT 1028

Qy   2062 GAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAG 2121
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 GAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAG 1088

Qy   1714 GGTGACAAGTGTCTGGACGGCGTGCTGAACAACAGCTACGAGTCAGACATCCTGAAG 1773
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4662 GGTGACAAGTGTCTGGACGGCGTGCTGAACAACAGCTACGAGTCAGACATCCTGAAG 4721

Qy   1774 AATTACCTGGCAACCAGAGAGTTTGACATGGGAAAAACATGTTGACCGAGAGCCTGGCT 1833
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4722 AATTACCTGGCAACCAGAGAGTTTGACATGGGAAAAACATGTTGACCGAGAGCCTCATGGCT 4781

Qy   1834 CTTCCAGCGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACCCGGACACCGTTCTGTGT 1893
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4782 CTTCCAGCGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACCCGGAGACCGTTCTGTGT 4841

Qy   1894 TACTGCTGTGGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAGTCAGCGTAACTGC 1953
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4842 TACTGCTGTGGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAGTCAGCAGCAACATTCCT 4901

Qy   1954 GCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCTGACTGCTACTGGGGCCGTAACTGC 2013
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4902 GCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCTGACTGCTACTGGGGCCGTAACTGC 4961

Qy   2014 CGCACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATATCTGTGAACAGACAAGG 2073
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4962 CGCACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATATCTGTGAACAGACAAGG 5021

Qy   2074 TTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAGGTGAAGAGACG 2133
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5022 TTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAGGTGAAGAGACG 5081

Qy   2134 TGTTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTTCACAGCCCCCTGAGGGGAAGGG 2193
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5082 TGTTTTTAAAATACAGAGACAACGCACGTCAAGGTGTTTTCACAGCCCCCTGAGGGGAAGGG 5141

Qy   2194 ACGCAGGTCTCCGACAGGTGCTCCTGGGGTGACTCTTCGTGGGAGCTTTTTACCCTCTGA 2253
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5142 ACGCAGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCGTGGGAGC--TTTACCCTCTGA 5199

Qy   2254 GTGAGACCCTCCCCAGAGCCCCGGGGGGCCCCGGGGGGCCCAGCCCGGCCCCGGTGGGCAG 2313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5200 GTGAGACCCTCCCCAGAGCCCCGGGGGGCCCCGGGGGGCCCAGCCCGGCCCCGGTGGGCAG 5259

Qy   2314 GGCTCGTGGTGGCATCAGCAGCAGAGACGAAGCCTTTCTGTAACATGCGGCCCGTCCCGCC 2373
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5260 GGCTCGTGGTGGCATCAGCAGCAGAGACGAAGCCTTTCTGTAACATGCGGCCCGTCCCGCC 5319

Qy   2374 GAGAGGGGCAGTTTGCTCTTTTGTACATTTTCCGAAACTACAGTTAAAGCAGAAGTCTG 2433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5320 GAGAGGGGCAGTTTGCTCTTTTGTACATTTTCCGAAACTACAGTTAAAGCAGAAGTCTG 5379

Qy   2434 TTTTCAGGAAAGTTTCAAGGGAGAAGGGCAAGTTTATCAAAAAACATTGTTTCAGGAGAA 2493
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5380 TTTTCAGGAAAGTTTCAAGGGAGAAGGGCAAGTTTATCAAAAAACATTGTTTCAGGAGAA 5439

Qy   2494 GGGAGCATAAGTTACAGCTACAGGGACGTACAACAATATCCTGCTGCTGGGAAAACCACA 2553
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5440 GGGAGCATAAGTTACAGCTACAGGGACGTACAACAATATCCTGCTGCTGGGAAAAACCACA 5499

Qy   2554 GCATTTTATCTATTTTTTATTTTAATAGGTTTGGTGCTTATAATTTATAATTTACAAATTGACAAAAT 2613
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5500 GCATTTTATCTATTTTTTATTTTAATAGGTTTGGTGCTTATCATTTATAATTTACAAATTGACAAAAT 5559

Qy   2614 GTCACAAACTGTAGCACAAATATATATAATTTATAATTTACAAATTGACAAAA 2666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5560 GTCACAAACTGTAGCACAAATATATATAATTTATAATTTACAAATTGACTAAA 5612

RESULT 15
AK090948
LOCUS   AK090948   2257 bp   mRNA   linear   PRI 15-JUL-2002
DEFINITION   Homo sapiens cDNA FLJ33629 fis, clone BRAMY2022454, highly similar
                to Homo sapiens cell cycle checkpoint protein CHFR mRNA.
ACCESSION   AK090948
VERSION   AK090948.1 GI:21749208
KEYWORDS   oligo capping; fis (full insert sequence).

```
Db   2519 TAGGATAGTCTCCTGTCTGCGAAAGACACAGCCTTTGTATATTTTTATTTTAATAAGTT 2578
               ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Qy   2585 TGGTGCTTATCTTCTAATAAGATTTAAAATGTCACAAACTGTAGCACAAATAATAATTT 2644
               |||||||||||||||||||||||||| |||||||| ||||||||||||| || |||||
Db   2579 TGGTGCTTATCTTCTAATAAGATTTAAATATTAAAAATGTAGCACAAATACTATAATTT 2638

Qy   2645 ATAATTTACAAATTGACAAAAAAAAAAAAA 2675
               ||||||||||||||| || |||
Db   2639 ATAATTTACAAATTGACTAAAATTAGGAATA 2669
```

RESULT 14

```
BD183291
LOCUS           BD183291                6235 bp    DNA     linear    PAT 17-JUN-2003
DEFINITION      Novel genes and proteins encoded by the genes.
ACCESSION       BD183291
VERSION         BD183291.1  GI:31875491
KEYWORDS        JP 2002345492-A/4.
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 6235)
  AUTHORS       Ohara,O., Nagase,T. and Nakajima,D.
  TITLE         Novel genes and proteins encoded by the genes
  JOURNAL       Patent: JP 2002345492-A 4 03-DEC-2002;
                KAZUSA DNA RESEARCH INSTITUTE
COMMENT         OS   Homo sapiens (human)
                PN   JP 2002345492-A/4
                PD   03-DEC-2002
                PF   26-FEB-2002 JP 2002049009
                PI   OSAMU OHARA,TAKAHIRO NAGASE,DAISUKE NAKAJIMA
                PC   C12N15/09,C07K14/47//A61K31/711,A61K38/00,A61K48/00,A61P25/00,
                PC   A61P25/14,
                PC   A61P25/18,A61P35/00,C12N15/00,A61K37/02
                CC   Novel genes and proteins encoded by the genes FH      Key
FT              Location/Qualifiers
                CDS             (50)..(1414).
                Location/Qualifiers
FEATURES        source          1..6235
                                /organism="Homo sapiens"
                                /mol_type="genomic DNA"
                                /db_xref="taxon:9606"
ORIGIN

Query Match           47.8%;  Score 1279.8;  DB 6;  Length 6235;
Best Local Similarity 99.5%;  Pred. No. 3.2e-283;
Matches 1306;  Conservative   0;  Mismatches   2;  Indels   5;  Gaps   2;
```

```
Qy   1354 AGCCAGCCATAGTCGTGTGCCGCAGTGTCCTGAGTACAGAAGGCAGCAGGCGGCGCAGCCT 1413
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4305 AGCCAGCCATACGTCGTGTGCCGCAGTGTCCTGAGTACAGAAGGCAGCAGGCGGCGCAGCCT 4364

Qy   1414 CCCCACTGCCCAGCACCCGAGGGCGAGCCAGGAGCCAGGAGCCCCTGGGGGGATGCACCC 1473
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4365 CCCCACTGCCCAGCACCCGAGGGCGAGCCAGGAGCCAGGAGCCCCTGGGGGGATGCACCC 4424

Qy   1474 TCCACGTCCGTCAGCGTCAGCGACAGCAGTCAGGATTACGTGTGCCCTGCAAGGAAGC 1533
               |||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db   4425 TCCACGTCCGTCAGCGTCAGCGACAGCAGTCAGGA---CAGTCCAGGATTACGTGTGCCCTGCAAGGAAGC 4481

Qy   1534 CACGCCCTGTGCACCTGCTTCCAGCCCATGCCCAGCCAGTGCGGTCGCGGAGAGCCGAGCAG 1593
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4482 CACGCCCTGTGCACCTGCTTCCAGCCCATGCCCAGCCAGTGCGGTCGCGGAGAGCCGAGCAG 4541

Qy   1594 GACCCGCGGTGTCGCCCCTCAGCAGTGTGCGGTCGGTCGCTGCTTCTGCAGCCTGTAC 1653
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4542 GACCCGCGGTGTCGCCCCTCAGCAGTGTGCGGTCGCTGCTTCTGCAGCCTGTAC 4601

Qy   1654 TGGGGCTGCACCCGGACCCGGACCCGGACCGCTCAACCTG 1713
               ||||||||||||||||||||||||||||||||||||||||
Db   4602 TGGGGCTGCACCCGGACCCGGACCCGGCTGCTCAACCTG 4661
```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="MGC:59419 IMAGE:6510480"
/tissue_type="Limb, mouse, day 10.5 and 11.5, pool of undifferentiated limb containing undifferentiated limb mesenchyme and early condensing mesenchyme."
/clone_lib="NIH_MGC_134"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6.1"

gene   1..3194
/gene="5730484M20Rik"
/note="synonym: C230082M18"
/db_xref="LocusID:231600"
/db_xref="MGI:1244898"

CDS   71..2065
/codon_start=1
/product="RIKEN cDNA 5730484M20"
/protein_id="AAH49792.1"
/db_xref="GI:29437111"
/db_xref="LocusID:231600"
/translation="MELHGBEQPPPQPWGRLLRLGAEBDEPQILLMKREWTIGRRR
GCDLSFPSNKLVSGDHCKLTVDEISGEVTLEDTSTNGTVINKLQVVKKQTYPLQSGDI
IVLVYRKNEPBHNVAYLYESLSGKQSLTQDSLEANKENMFHVTKDCSSPGQGDDPQVP
LLSPMAQTCLEEPQPSTSTSDLLPTASTSSTEPELTSAGQKHSSSSGPGNTSISPKGR
SSIVANGELSSLSPVFQDKEASFSLLESKDHEBLEPAKKKMKGDGELDTNLQLLVSGQ
RGNAQTSSEDVKDASVKPDKMEBTLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMER
SSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVRSMDARNKITQDMLQPK
VRRSFSDEBGSSEDLLELSDVDSESSSDISQPYIVCRQCPEYRRQAVQSLPCPVPESEL
GATIALGGBAPSTSASLPTAAPDYMCPLQGSHAICTCCFQPMDRRARREQDPRVAPQ
QCAVCLQPFCHLYWGCRTGCFGCLAPFCELNLGDKCLDGVLNNNYBSDILKNYLAT
RGLTWKSVLTESLLALQRGVFMLSDYRITGNTVLCYCCGLRSFRELTYQYRQNIPASE
LPVTVTSRPDCTWGRNCRTQVKAHHAMKFNHICBQTRPKN"

misc_feature 182..379
/note="FHA; Region: FHA domain. The FHA (Forkhead-associated) domain is a phosphopeptide binding motif"
/db_xref="CDD:pfam00498"

misc_feature 974..1105
/note="HRD1; Region: COG5243, HRD1, HRD ubiquitin ligase complex, ER membrane component [Posttranslational modification, protein turnover, chaperones]"
/db_xref="CDD:COG5243"

ORIGIN

Query Match   55.7%; Score 1491; DB 10; Length 3194;
Best Local Similarity 75.2%; Pred. No. 0;
Matches 2008; Conservative 0; Mismatches 615; Indels 48; Gaps 10;

Qy   2665 AA 2666
Db   1501 AA 1502

RESULT 13
BC049792
LOCUS   BC049792   3194 bp   mRNA   linear   ROD 07-OCT-2003
DEFINITION Mus musculus RIKEN cDNA 5730484M20 gene, mRNA (cDNA clone MGC:59419
            IMAGE:6510480), complete cds.
ACCESSION   BC049792
VERSION   BC049792.1 GI:29437110
KEYWORDS   MGC.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 3194)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grinwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2 (bases 1 to 3194)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (31-MAR-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@cgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 109 Row: l Column: 17
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 27370049.
FEATURES   Location/Qualifiers
     source   1..3194

Qy   1585 CGCGAGCAGGACCCGCGTGTGCGCCCTCAGCAGTGCGGCTCGCCTGCCAGCCTTTCTGC 1644
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CGCGAGCAGGACCCGCGTGTGCGCCCTCAGCAGTGCGGCTCGCCTGCCAGCCTTTCTGC 480

Qy   1645 CACCTGTACTGGGGCTGCGCACCCGGACCGGCTGCTACGGCGCTGCGGCTGGCCCCGTTTTGTGAG 1704
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CACCTGTACTGGGGCTGCGCACCCGGACCGGCTGCTACGGCGCTGCGGCTGGCCCCGTTTTGTGAG 540

Qy   1705 CTCAACCTGGGTGACAAGTGTCTGGACGGCGTGCTGAACAACAGCTACGAGTCAGAC 1764
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CTCAACCTGGGTGACAAGTGTCTGGACGGCGTGCTGAACAACAGCTACGAGTCAGAC 600

Qy   1765 ATCCTGAAGAATTACCTGGCAACCAGAGAGGTTGACATGGAAAAACATGTTGACCGAGAGC 1824
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ATCCTGAAGAATTACCTGGCAACCAGAGAGGTTGACATGGAAAAACATGTTGACCGAGAGC 660

Qy   1825 CTCGGTGGCTCTCCAGCGGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACACC 1884
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CTCGGTGGCTCTCCCAGCGGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGCGGAGACACC 720

Qy   1885 GTTCGTGTTACTGCTGTGTGCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCGAG 1944
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GTTCGTGTTACTGCTGTGTGCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCGAG 780

Qy   1945 AACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCGTCCGACTGCTACTGCGGGGC 2004
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 AACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCGTCCGACTGCTACTGCGGGGC 840

Qy   2005 CGTAACTGCCCACTCAGGTGAAAGCTCACACCGGCCATGAAATTCAATCATATCTGTGAA 2064
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CGTAACTGCCGCACCAGGTGAAAGCTCACACCGGCCATGAAATTCAATCATATCTGTGAA 900

Qy   2065 CAGACAAGGTTCAAAAACTAAGCATCCAGAGCCCTGAGCAGTTTCAGCACTGGAGGTG 2124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 CAGACAAGGTTCAAAAACTAAGCATCCAGAGCCCTGAGCAGTTTCAGCACTGGAGGTG 960

Qy   2125 AAGGAGAGCGTGTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTTCACAGCCCCCTG 2184
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAGGAGAGCGTGTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTTCACAGCCCCCTG 1020

Qy   2185 AGGGAAGGGACGCAGGGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCTGTGGGAGCTTTTT 2244
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGGGGAAGGGGACGCAGGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCTGTGGGAGCTTTTT 1080

Qy   2245 ACCCTCTGAGTGAGACCCCTCCCCAGAGCCCCGGGGCCCCAGCCCCCTCCGTGGTGAGC 2304
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ACCCTCTGAGTGAGACCCCTCCCCAGAGCCCCGGGGCCCCAGCCCCCTCCGTGGTGAGC 1140

Qy   2305 GCTGGGCAGGGCTCGTGGTGCCATCAGCAGCAGAGACGAGAGCCTTCTGTAACATGCGGC 2354
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GCTGGGCAGGGCTCGTGGTGCCATCAGCAGCAGAGACGAGAGCCTTCTGTAACATGCGGC 1200

Qy   2365 CGTCCCGCCAGAGAGGGGGCAGTTTGCTCTTTGTACATTTTCCGAAACTACAGTTAAAGC 2424
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CGTCCCGCCAGAGAGGGGGCAGTTTGCTCTTTGTACATTTTCCGAAACTACAGTTAAAGC 1260

Qy   2425 AGAAGTCGTGTTTTCAGGAAAAGTTTCAGGAGAGGGGCAAGTTTATCAAAAACATTGTT 2484
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AGAAGTCTGTTTTCAGGAAAAGTTTCAGGAGAGGGGCAAGTTTATCAAAAACATTGTT 1320

Qy   2485 TCAGGAGAAGGGAGCCATAAGTTTACAGCCTACACAGGAGCGTACACAATAATCCTGCTGCTGG 2544
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TCAGGAGAAGGGAGCCATAAGTTTACAGCCTACACAGGAGCGTACACAATAATCCTGCTGCTGG 1380

Qy   2545 AAAACCACAGCAGCATTTTATCATTTTTATTTTAATAGGTTGGTTGCTTAATCTTCTAATAA 2604
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AAAACCACAGCAGCATTTTATCATTTTTATTTTAATAGGTTGGTTGCTTAATCTTCTAATAA 1440

Qy   2605 GATTTAAATGTCACAAACTGTAGCACAAATAATAATATTATAATTACAAATTGACAAA 2664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GATTTAAATGTCACAAACTGTAGCACAAATAATAATATTATAATTACAAATTGACTAA 1500

COMMENT   Martinsried, GERMANY
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. This clone
            (DKFZp434N2420) is available at the RZPD in Berlin. Please contact
            the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES   Location/Qualifiers
     source   1..2297
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434N2420"
                     /tissue_type="testis"
                     /clone_lib="434 (synonym: htes3). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
     gene   1..2297
                     /gene="DKFZp434N2420"
     CDS   <1..921
                     /gene="DKFZp434N2420"
                     /note="unknown"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB70812.1"
                     /db_xref="GI:6808266"
                     /db_xref="SPTREMBL:Q9NT32"
                     /translation="VEAYLIOHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEE
                     GSSEDLELSDVDSESSDISOPYVVCRQCPEYRRQAAQPHCPAPEGEPGAPQALGDA
                     PSTSVSLTTAVQDYVCPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFC
                     HLYWGCTRTGCYGCLAPFCELNLGDKCLDGVLNNSYESDIIKNYLATRGLTWKNMLT
                     BSLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQRQNIPASELPVATSRPD
                     CYWGRNCRTQVKAHHAMKPNHICEQIRFKN"
     polyA_signal   2227..2232
                     /gene="DKFZp434N2420"
     polyA_site   2251
                     /gene="DKFZp434N2420"

ORIGIN

Query Match   56.0%;   Score 1500.4;   DB 9;   Length 2297;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches 1501;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1165 GTGGAAGCATACCTCATCCAGCATCCAGACAACAAGAGTCGCAGTGAAGAAGATGTGCAAAGT 1224
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTGGAAGCATACCTCATCCAGCATCCAGACAACAAGAGTCGCAGTGAAGAAGATGTGCAAAGT 60

Qy   1225 ATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCGAGCCCAAAGTCAGGCCGGTCTTTT 1284
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCGAGCCCAAAGTCAGGCCGGTCTTTT 120

Qy   1285 TCTGATGAAGAAGGAGTTCAGGAGGACCTGCTGAGCGTGCAGACGTTGACAGTGAGTCC 1344
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TCTGATGAAGAAGGAGTTCAGGAGGACCTGCTGAGCGTGCAGACGTTGACAGTGAGTCC 180

Qy   1345 TCAGACATTAGCCAGCCATACGTCGTGTCCGGGCAGTGTCCTGAGTACGAAGGCAGGCG 1404
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCAGACATTAGCCAGCCATACGTCGTGTCCGGGCAGTGTCCTGAGTACGAAGGCAGGCG 240

Qy   1405 GCGCAGCCTCCCACTGCCCAGCACCCGAGGGCGAGCCAGGAGCCCCACAGGCCCTGGGG 1464
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCGCAGCCTCCCACTGCCCAGCACCCGAGGGCGAGCCAGGAGCCCCACAGGCCCTGGGG 300

Qy   1465 GATGCACCCTCCAGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTGTGCCCTCTG 1524
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GATGCACCCTCCAGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTGTGCCCTCTG 360

Qy   1525 CAAGGAAGCCACGCCCTGTGCACCTGCTGCTTCCAGCCCATGCCCAGCCCGAGAGCGGAG 1584
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CAAGGAAGCCACGCCCTGTGCACCTGCTGCTTCCAGCCCATGCCCAGCCCGACCGGAGAGCGGAG 420

Db   1547 GGCTCTCCAGCGGCGGGAGTGTTTCTGCTGTCTGATTACAGAGATCACGGGAGACACCGTTCT 1606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1890 GTGTTACTGCTGTGGCCTGCAGCTTCGGTGAGCTGACCTATCAGTATCGGCAGAAACAT 1949
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1607 GTGTTACTGCTGTGGCCTGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACAT 1666

Qy   1950 TCCTGCTTCCGAGTGCCAGTGGCCGTAACATCCGTCCGACTGCTACTGGGGCCGTAA 2009
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1667 TCCTGCTTCCGAGTGCCAGTGGCCGTAACATCCGTCCGACTGCTACTGGGCCGTAA 1726

Qy   2010 CTGCCCACTCAGGTGAAAGCTCACCACGCCCATGAAATTCAATCATATCTGTGAACAGAC 2069
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1727 CTGCCCACTCAGGTGAAAGCTCACCACGCCCATGAAATTCAATCATATCTGTGAACAGAC 1786

Qy   2070 AAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTCAGCACTGAGGTGAAGAG 2129
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1787 AAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTCAGCACTGAGGTGAAGAG 1846

Qy   2130 AGCGTGTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTCACAGCGCCCCTGAGGGA 2189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1847 AGCGTGTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTCACAGCGCCCCTGAGGGA 1906

Qy   2190 AGGGACGCGCAGGGTCTCCGACAGGTGCTGCTGGGGTGACTCTTCTGTGGAGCTTTTACCCT 2249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1907 AGGGACGCGCAGGGTCTCCGACAGGTGCTGCTGGGGTGACTCTTCTGTGTGGAGCTTTTACCCT 1966

Qy   2250 CTGAGTGAGTGAGACCCCTCCCAGAGCCCCGAGAGCCCCGGGGGCCGCAGCCCGAGCCGCTG 2309
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1967 CTGAGTGAGTGAGACCCCTCCCAGAGCCCCAGAGCCCCGGGGGCCGCAGCCCGAGCCGCTG 2026

Qy   2310 GCAGGGCTCGTGGTGGCATCAGCAGCAGCAGAGACGAAGCCTTTCTGTAACATGCGGCCGTCC 2369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2027 GCAGGGCTCATGGTGGCATCAGCAGCAGCAGAGACGAAGCCTTTCTGTAACATGCGGCCGTCC 2086

Qy   2370 CGCCCGAGAGGGCAGTTTGCTCTTTTGTACATTTTCCGAAACTACAGTTAAAGCACAGAAG 2429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2087 CGCCCGAGAGGGCAGTTTGCTCTTTTGTACATTTTCCGAAACTACAGTTAAAGCACAGAAG 2146

Qy   2430 TCTGTTTTCAGGAAAAGTTCAAGGGAGAAGGGGCAAGTTTATCAAAAACATTGTTTCAGG 2489
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2147 TCTGTTTTCAGGAAAAGTTCAAGGGAGAAGGGGCAAGTTTATCAAAAACATTGTTTCAGG 2206

Qy   2490 AGAAGGGAGCATAAAGTTTACAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAAC 2549
          ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db   2207 AGAAGGGAGCATAAAGTTTACAGCCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAAC 2266

Qy   2550 CACAGCATTTATCTATTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATT 2609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2267 CACAGCATTTATCTATTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATT 2326

Qy   2610 AAATGTCACAAACTGTAGCACAAATAATAATTATAATTTACAAATTGACAAAA 2666
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   2327 AAATGTCACAAACTGTAGCACAAATAATAATTATAATTTACAAATTGACTAAAA 2383

RESULT 12
HSM802304
LOCUS   HSM802304   2297 bp   mRNA   linear   PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA; cDNA DKFZp434N2420 (from clone DKFZp434N2420);
            partial cds.
ACCESSION   AL137561
VERSION   AL137561.1   GI:6808265
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2297)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152

```
                Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and
                Isogai,T.
  TITLE         NEDO human cDNA sequencing project
  JOURNAL       Unpublished
REFERENCE     2  (bases 1 to 2448)
  AUTHORS       Isogai,T. and Yamamoto,J.
  TITLE         Direct Submission
  JOURNAL       Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
                Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
                (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT       NEDO human cDNA sequencing project supported by Ministry of
                Economy, Trade and Industry of Japan; cDNA full insert sequencing:
                Research Association for Biotechnology (RAB); cDNA library
                construction: Helix Research Institute (HRI) (supported by Japan
                Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
                HRI, and Biotechnology Center, National Institute of Technology and
                Evaluation; clone selection for full insert sequencing: HRI and RAB.
                RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
  source          1..2448
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="TESTI2033505"
                  /tissue_type="testis"
                  /clone_lib="TESTI2"
                  /note="cloning vector: pME18SFL3"
ORIGIN
Query Match      71.6%;  Score 1919.4;  DB 9;  Length 2448;
Best Local Similarity  88.5%;  Pred. No. 0;
Matches 2227;  Conservative  0;  Mismatches  6;  Indels 284;  Gaps  2;
```

(DNA sequence alignment data — Query (Qy) and Database (Db) aligned sequences with position numbers)

RESULT 11
LOCUS   AK097671   2448 bp   mRNA   linear   PRI 15-JUL-2002
DEFINITION   Homo sapiens cDNA FLJ40352 fis, clone TESTI2033505, highly similar
               to Homo sapiens cell cycle checkpoint protein CHFR mRNA.
ACCESSION   AK097671
VERSION   AK097671.1 GI:21757518
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Butheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
               Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
               Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
               Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
               Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
               Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,

```
RESULT 10
AX834949
LOCUS           AX834949                2448 bp    DNA             linear   PAT 15-DEC-2003
DEFINITION      Sequence 2073 from Patent EP1347046.
ACCESSION       AX834949
VERSION         AX834949.1  GI:39921084
KEYWORDS        .
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1
  AUTHORS       Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
                Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
                Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
                Masuho,Y.
  TITLE         Full-length cDNA sequences
  JOURNAL       Patent: EP 1347046-A 2073 24-SEP-2003;
                Research Association for Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1..2448
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
Query Match          71.6%;  Score 1919.4;  DB 6;  Length 2448;
Best Local Similarity   88.5%;  Conservative   0;  Indels  284;  Gaps   2;
Matches 2227;          Mismatches   6;    Conservative   0;
```

```
Qy    218 GACGAGGTTGCGACCTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAA 277
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db     83 GACCCTGTTGCGACCTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAA 142

Qy    278 TTGTAGTGATGAAAAATCAGGTCAGGTGACACTGGAAGATACCAGCACCAGTGGAACAG 337
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    143 TTGTAGTGATGAAAAATCAGGTCAGGTGACACTGGAAGATACCAGCACCAGTGGAACAG 202

Qy    338 TGATTAACAAGCTGAAGGTTGTTAAGAGCAGCACAGCATGCCCTTTACAGACTGGGGATGTCA 397
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    203 TGATTAACAAGCTGAAGGTTGTTAAGAGCAGCACAGCATGCCCTTTACAGACTGGGGATGTCA 262

Qy    398 TCTACTTGGTGTACAGGAAGAATGAACCGGAACACGTGGCATACCTCTCATGAAATCTT 457
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    263 TCTACTTGGTGTACAGGAAGAATGAACCGGAACACGTGGCATACCTCTCATGAAATCTT 322

Qy    458 TAAGTGAAAAGCAAGGCATGACAACAAGAATCTTTGAAGCTAACAAGGAAAATGTGTTCC 517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    323 TAAGTGAAAAGCAAGGCATGACAACAAGAATCCTT------------------------- 356

Qy    518 ATGGGACCAAAGATACCTTCAGGTGCAGGTGCAGGGGCCGATCCCCGGGTCCCTC 577
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357 --------------------------------------------------------- 356

Db    578 CGTCGTCGCCGCGCCACTCCAGCTGTGCTTTGAGGAACCACAGCCATCAACATCGACGTCAG 637

Db    357 --------------------------------------------------------- 356

Qy    638 ACCTCTTCCCCACAGCCTCGGCCTCTTCCCAGCTCTTCCAGCGGGGCGAGAGCGTT 697

Db    357 --------------------------------------------------------- 356

Qy    698 CCTCCAGTGTCTGGGCTCTGGGGGTGGTGGCCATCTCCCCTAAAGGAAGTTGGTCCCTCCTGTGG 757

Db    357 -----------TGGGTCTGGGGGTGGTGGCCATCTCCCCTAAAGGAAGTTGGTCCCTCTGTGG 406

Qy    758 CAAGTGATGAAGTCTCCAGCTTTGCCTCAGCTCTTCCAGCTCTCTTT 817
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    407 CAAGTGATGAAGTCTCCAGCTTTGCCTCAGCTTTGCCTCAGCTCTCTTT 466

Qy    818 CGTCGTTGAACCCCAGGATCAGGAGATTTGGAGCCCGTGAAGAAAAATGAGAGGAG 877
```

```
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Tang,Y.T., Liu,C., Zhou,P., Asundi,V., Zhang,J., Zhao,Q.A., Ren,F.,
            Xue,A.J., Yang,Y., Wehrman,T. and Drmanac,R.T.
TITLE       Novel nucleic acids and polypeptides
JOURNAL     Patent: WO 0222660-A 57 21-MAR-2002;
            HYSEQ, INC. (US)
FEATURES    Location/Qualifiers
source      1..3181
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
CDS         96..1967
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAD34749.1"
            /db_xref="GI:21438724"
            /db_xref="REMTREMBL:CAD34749"
            /translation="MERPEEGKQSPPPQPWGRLLRLGAEBEGEPHVLLRKREWTIGRRR
            GCDLSFPSNKLVSGDHCRIVVDBKSGGVTLEDTSTSGVVINKLKVVKKQTCPLQTGDV
            IYLVYRKNEPEHNVAYLYESLSEKQGMTVQESFEMVPCCVAQAGLKLJGSSDPPTLASQ
            SIVITGSGGGISPKGSGSPSVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPVKKK
            MRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMBETLTCIICQDLLHDCVSL
            QPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEBED
            VQSMDARNKITQDMLQPKVRRSFSDEBGSSEDLELSDVDSESSDISQPYYVCRQCPE
            YRRQAAQPPHCPAPEGEPGAPQALGDAPPTSVSLTTAVQDYVCPLQQSHALCTCCFQP
            MPDRRABREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELNLGDKCLDGV
            LNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLR
            SFRELTTQYQQNIPASELPVAVTSRPDCYWGRNCRTQVKAHAMKFNHICBQTRFKN"
ORIGIN

Query Match    83.8%;   Score 2244.6;   DB 6;   Length 3181;
Best Local Similarity  92.3%;   Pred. No. 0;
Matches 2447;   Conservative  0;   Mismatches  79;   Indels 125;   Gaps  3;
```

```
Qy    16 GGGCCGCAAATGTCTCTTGACAGCGGCGGCGGCCGGCGGCGGTTCGGGCGTTCGGGGCG 75
         |||||||  ||||||||||||||||||||||||||||| |||||||||||||||||||
Db    21 GGCCCGTACCGGTCCGGAATTCCCGGGTCCGACGATTTCGTTTCGGGCGTTCGGGGGCG 80

Qy    76 GGGATGTGAATCCCGATGGAGCGGCCCGAGGAAGGCAAGCAAGCAGTCGCCGCCGCCCC 135
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    81 GGGATGTGAATCCCGATGGAGCGGCCCGAGGAAGGCAAGCAAGCAGTCGCCGCCGCCCC 140

Qy   136 TGGGGACGGCGGCTCCTGCGTCTGGGCGGCGGAGGAGGGGCACGTCCTCCTGAGGAAG 195
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 TGGGGACGGCGGCTCCTGCGTCTGGGCGGCGGAGGAGGGGCACGTCCTCCTGAGGAAG 200

Qy   196 CGGGAGTGGACCATCGGGCGGAGACGAGGTTGCGACCTTTCCTTCCGCAGCAATAAACTG 255
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201 CGGGAGTGGACCATCGGGCGGAGACGAGGTTGCGACCTTTCCTTCCGCAGCAATAAACTG 260

Qy   256 GTCTCTTGGAGATCACTGTAGAAATTGTAGTGGATGAATGAAAATCAGGTGACACTGGAA 315
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 GTCTCTTGGAGATCACTGTAGAAATTGTAGTGGATGAATGAAAATCAGGTGACACTGGAA 320

Qy   316 GATACCACGACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGACATGC 375
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 GATACCACGACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGACATGC 380

Qy   376 CCTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAATGAACCGGAACACAAC 435
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   381 CCTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAATGAACCGGAACACAAC 440

Qy   436 GTGGCATACCTCTATGAATCTTAAGTGAAAAGCAAGGCATGAGACACAAGAATCCTTTGAA 495
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   441 GTGGCATACCTCTATGAATCTTAAGTGAAAAGCAAGGCATGAGACACAAGAATCCTTTGAG 500

Qy   496 GCTAACAAGGAAAATGTGTTCCATGGGACCCAAAGATACCTTCAGGTGCAGGGGCGA 555
         |||    ||||||       |||||                ||||
```

```
Db   501 ATGTGCCTTGCTGTCGTGTTGCCCAGGCTGGTCGTCCGGGTCGGGTGTGCTTGAAAACTTCCTGCTCAAAACTCCTCGGATCAAGT-------- 551
Qy   556 GGGGCCGATCCCGGGGTCCTCCGTCCGTCCGGGTCGGGTGTGCTTTGGCCCTCCCAAAGTATTGTGATTACA------- 615
Db   552 -------------GATCTCCACCTTGGCCCTCCCAAAGTATTGTGATTACA------- 590
Qy   616 CAGCCATCAACATCGACGTCAGACCTCTTCCCCACAGCCTCGGCCTCTTCCACGGGAGCCT 675
Db   591 ------------------------------------------------------------ 590
Qy   676 TCTCCTGCAGGGCCAGAGCGTTCCTCCAGTGTGGGTCTGGGGGTGTGGGCATCTCCCCT 735
Db   591 -----------------------------GGGTCTGGGGGTGTGGGCATCTCCCCT 617
Qy   736 AAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTGCCTCAGCTCTCCCA 795
Db   618 AAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTGCCTCAGCTCTCCCA 677
Qy   796 GACACAGAAGACTCGGCTCCTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTTGGAGCC 855
Db   678 GACAGAAGACTCGGCTCCTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTTGGAGCC 737
Qy   856 GTGAAGAAGAAAATGAGAGGAGATGGGGACCTGACCTGAAGCGGCAGTTGTTGGTCGCA 915
Db   738 GTGAAGAAGAAAATGAGAGGAGAGATGGGGACCTTGACCTGAAGCGGCAGTTGTTGGTCGCA 797
Qy   916 CAACCGGTAGAAATGCCCAAACGTCACGAGGACGTCAGAGCAGCAGCGGCGGCTGGGAAGCCA 975
Db   798 CAACCGGTAGAAATGCCCAAACGTCACGAGGACGTCAGAGCAGCAGCGGCGGCTGGGAAGCCA 857
Qy   976 GACAAGATGGAGAGGACGCTGACATGCATCATCTGCCAGGACCTGCTGCACGACTGCGTG 1035
Db   858 GACAAGATGGAGAGGACGCTGACATGCATCATCTGCCAGGACCTGCTGCACGACTGCGTG 917
Qy  1036 AGTTTGCAGCCCCTGCCATGCACAGTTCTGCGCGGCCTTGCTACTCGGGCTGGATGGAGCGC 1095
Db   918 AGTTTGCAGCCCCTGCCATGCACAGTTCTGCGCGGCCTTGCTACTCGGGCTGGATGGAGCGC 977
Qy  1096 TCGTCCCTGTGTCCTACCTGCCGGATCGTGAGCGGGATCGTGAAAAACCACATCCTC 1155
Db   978 TCGTCCCTGTGTCCTACCTGCCGGATCGTGAGCGGGATCGTGAAAAACCACATCCTC 1037
Qy  1156 AACAACCTCGTGAAGCAGCATAGCCAGACATCCAGACAAGAGTCGCAGTGAAGAAGAGAT 1215
Db  1038 AACAACCTCGTGAAGCAGCATAGCCAGACATCCAGACAAGAGTCGCAGTGAAGAAGAGAT 1097
Qy  1216 GTGCAAAGTATGGATGCCAAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGG 1275
Db  1098 GTGCAAAGTATGGATGCCAAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGG 1157
Qy  1276 CGGTCTTTTTCTGATGAAGAAGGGAGTTCAGAGAACCTGCTGGAGCTGTCAGACGTTGAC 1335
Db  1158 CGGTCTTTTTCTGATGAAGAAGGGAGTTCAGAGAACCTGCTGGAGCTGTCAGACGTTGAC 1217
Qy  1336 AGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTCCGGCAGTGTCCTGAGTACAGA 1395
Db  1218 AGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTCCGGCAGTGTCCTGAGTACAGA 1277
Qy  1396 AGGCAGGCGGCGGCCCTCCCCACTGCCCCGAGGGCGGCAAGGAGCCCCACAG 1455
Db  1278 AGGCAGGCGGCGGCCCTCCCCACTGCCCCGAGGGCGGCAAGGAGCCCCACAG 1337
Qy  1456 GCCCTGGGGATCCACCCTTCCAGTGCCTCAGCCTGACGACAGCAGCAGTCCAGGATTACGTG 1515
Db  1338 GCCCTGGGGGATGCACCCTTCCAGTGCCTCAGCCTGACGACAGCAGCAGTCCAGGATTACGTG 1397
Qy  1516 TGCCCTCTGCAAGGAAGCCACGCCCTGCTTGCACCTGCTTCCAGCCATGCCCATGCCACCGG 1575
Db  1398 TGCCCTCTGCAAGGAAGCCACGCCCTGCTTGCACCTGCTTCCAGCCATGCCCATGCCACCGG 1457
Qy  1576 AGAGCGGAGCGCCGAGCAGGACCCGCGTGTCGCCCTCAGCAGTGCGGTCTGCCTGCAG 1635
Db  1458 AGAGCGGAGCGCCGAGCAGGACCCGCGTGTCGCCCTCAGCAGTGCGGTCTGCCTGCAG 1517
```

RESULT 9
AX405642
LOCUS   AX405642   3181 bp   DNA   linear   PAT 14-JUN-2002
DEFINITION   Sequence 57 from Patent WO0222660.
ACCESSION   AX405642
VERSION   AX405642.1   GI:21438723

```
                                 etc.) and Department of Virology, Institute of Medical Science,
                                 University of Tokyo.
FEATURES             Location/Qualifiers
     source          1..3138
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RP4000648"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2RP4"
                     /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
     CDS             65..1936
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA91817.1"
                     /db_xref="GI:7023051"
                     /translation="MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRR
                     GCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGVINKLKVVKKQTCPLQTGDV
                     IYLVYRKNEPEHNVAYLYESLSEKQGMTQESFEMVPCCVAQAGLKLLGSSDPPTLASQ
                     SIVITGSGSGGISPKGSGPSVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPEKKK
                     MRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEETLTCIICQDLLHDCVSL
                     QPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEED
                     VQSMDARNKITQDMLQPKVRRSFSDEGSSEDLLELSDVDSSSSDISQPYVVCRQCPE
                     YRRQAAQPPHCPAPEGEPGAPQALGDAPPTSVSLTTAVQDYVCPLQGSHALCTCCQFP
                     MPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELNLGDKCLDGV
                     LNNSYESDILKNYLAIRGLTWKQMLITESLVALQRGVFLLSDYRVTGDTVLCYCCGLR
                     SFRELTYQYQQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQTRFKN"
ORIGIN

Query Match             84.4%;  Score 2262.4;  DB 9;  Length 3138;
Best Local Similarity   93.0%;  Pred. No. 0;
Matches  2454;  Conservative   0;  Mismatches   0;  Indels 125;  Gaps   3;

Qy   27  CTCTTGACAGCGGCGGCGGCCGCCGGTTCGGCGGCCGGGGCGGGGGGGGGATGTGAAT    86
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  CTCTTGACAGCGGCGGCGGCCGCCGGTTCGGCGGCCGGGGCGGGGGGGGGATGTGAAT    60

Qy   87  CCCGATGGAGCGGGCCCGAGAAGGAGCAAGCAGTCGCCGCCACGTCGCCCGGGACGGGCT   146
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  CCCGATGGAGCGGGCCCGAGAAGGAGCAAGCAGTCGCCGCCACGTCGCCCGGGACGGGCT   120

Qy  147  CCTGCCTCTGGCGCGGAGGAGGGCGAGCCGCACGTCCTCCTGAGGAAGCGGGAGCGGTGGAC   206
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CCTGCCTCTGGCGCGGAGGAGGGCGAGCCGCACGTCCTCCTGAGGAAGCGGGAGCGGGAC    180

Qy  207  CATCGGGCGGAGACGAGGTTGCGACCTTTCCTCCCCAGCAATAAACTGGTCTCTGGAGA    266
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  CATCGGGCGGAGACGAGGTTGCGACCTTTCCTCCCCAGCAATAAACTGGTCTCTGGAGA    240

Qy  267  TCACTGTAGAATTGTAGTGGATGAGAAAATCAGGTCAGGTGACACTGACACTGGAAGATACCAGCAC   326
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  TCACTGTAGAATTGTAGTGGATGAGAAAATCAGGTCAGGTGACACTGACACTGGAAGATACCAGCAC   300

Qy  327  CAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGCAGCATGCCCTTTACAGAC   386
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  CAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGCAGCATGCCCTTTACAGAC   360

Qy  387  TGGGGATGTGCATCTCACTTGGTGTACAGGAAGAATGAACCGGAACAACCAACGTGGCATACCT   446
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  TGGGGATGTGCATCTCACTTGGTGTACAGGAAGAATGAACCGGAACAACCAACGTGGCATACCT   420

Qy  447  CTATGAATCTTTAAGTGAAAAGCAAGGCATGACACAAGAATCCTTTGAAGCTAACAAGGA    506
         |||||||||||||||||||||||||||||||||||||||||
Db  421  CTATGAATCTTTAAGTGAAAAGCAAGGCATGACACAAGAATCCTTTGAGATGGTGCCTTG    480

Qy  507  AAATGTGTTCCATGGACCAAAGATACCTCAGGTGCAGGGGCGAGGGGCCGATCC         566
         ||||||||||||||||||||||||||
Db  481  CTGTGTTGCCCAGGTGGTCGGTCTAAAACTCCTGGGATCAAGT-----------        520

Qy  567  CCGGGTCCCTCCGTCGTGCCGCCCACTCAGGTGCTTTGAGGAACCAGCCATCAAC         626
```

```
Db 1858  TAACTGCCGCACTCAGTGAAAGCTCACCACGCCATGAAATTCAATCATATCTGTGAACA   1917

Qy 2067  GACAAGGTTCAAAAACTAAGCATCCAGAGAGGGCCTGAGCAGTTTCAGCACTGGAGGTGAA   2126
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1918  GACAAGGTTCAAAAACTAAGCATCCAGAGAGGGCCTGAGCAGTTTCAGCACTGGAGGTGAA   1977

Qy 2127  GAGAGGCGTGTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTTCACAGCCCCCTGAG   2186
         ||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||
Db 1978  GAGAGGCGTGTTTTTAAAATACAGAGGCAAGCACGTCAAGGTGTTTTCACAGCCCCCTGAG   2037

Qy 2187  GGAAGGGACGCAGGGTCCCGACAGGTCCTCGGGTGACTCTTCGTGTGGAGCTTTTTAC     2246
         |||||||||||||| ||| ||||||| ||||||||||||||| |||||||   ||||
Db 2038  GGAAGGGACGCAGGGGTCTCCGACAGGTGCTCGGGTGACTCTTCTGTGGAGC--TTTAC     2095

Qy 2247  CCTCTGAGTGAGACCCTCCCCAGAGCCCCGCAGCCCGCCCTCCTGGTGAGCGC         2306
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2096  CCTCTGAGTGAGACCCTCCCCAGAGCCCCGCAGCCCGCCCTCCTGGTGAGCGC         2155

Qy 2307  TGGGCAGGGCTCGTGGTGGCATCAGCAGCAGAGACGAGAGCCTTTCTGTAACATGCCGCCG   2366
         |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db 2156  TGGGCAGGGCTCGTGGTGGCATCAGCAGCAGAGAAGCCTTTCTGTAACATGCGCGCCG     2215

Qy 2367  TCCCGCCGAGAGGGGGCAGTTTTGCTCTTTGTACATTTTCCGAAACTACAGTTAAAGCAG   2426
         ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db 2216  TCCTGCCGAGAGGGGGCAGTTTTGCTCTTTGTACATTTTCCGAAACTACAGTTAAAGCGG   2275

Qy 2427  AAGTCTGTTTTCAGGAAAAGTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTC   2486
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2276  AAGTCTGTTTTCAGGAAAAGTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTC   2335

Qy 2487  AGGAGAAGGGAGCATAAGTTTACAGCCTACAGGAGCTACACAATATCCTGCTGCTGGGAA   2546
         |||||||||| ||||||||||||||||||||||| |||||||||||||||||||||||||
Db 2336  AGGAGAAGGGGAGCATAAGTTTACAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAA   2395

Qy 2547  AACCACAGCATTTATCTCATTTTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGA   2606
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2396  AACCACAGCATTTATCTCATTTTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGA   2455

Qy 2607  TTTAAATGTCACAAACTGTAGCACAAATAATAATAATTATATTTATAATTACAAATTGACAAAAA   2666
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db 2456  TTTAAATGTCACAAACTGTAGCACAAATAATAATAATTATATTTATAATTACAAATTGACTAAAA   2515
```

```
RESULT 8
AK001658
LOCUS       AK001658                3138 bp    mRNA    linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ10796 fis, clone NT2RP4000648, weakly similar
            to TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP0.
ACCESSION   AK001658
VERSION     AK001658.1  GI:7023050
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3138)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,
            Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
            Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3138)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
```

```
FEATURES             Location/Qualifiers
FT     CDS
       source          1..3138
                       /organism="Homo sapiens"
                       /mol_type="genomic DNA"
                       /db_xref="taxon:9606"
       CDS             (65)..(1933)

ORIGIN

Query Match      84.4%;   Score 2262.4;   DB 6;   Length 3138;
Best Local Similarity   93.0%;   Pred. No. 0;
Matches 2454;   Conservative 0;   Mismatches 0;   Indels 61;   Gaps 3;

Qy     27  CTCTTGACACGCGGCGGCGGCGGGGCCCAGCGGTTCGGCGTCGGCGCGGGGCGGGGATGTGAAT   86
           |||||||  ||||||||||||||||  |||||||||||||||||||||||||||||
Db      1  CTCTTGACAGCGGCGGCGGCGGGGTTCGGGTTCGGCGCGGGGCGGGGATGTGAAT          60

Qy     87  CCGATGGAGCGGCCCGAGGAAGCAAGAGTCGCCGGCGCCCGCAGCCCTGGGGACGGCT       146
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  CCGATGGAGCGGCCCGAGGAAGGCAAGCAGTCGCCGGCGCCCGCAGCCCTGGGGACGGCT     120

Qy    147  CCTGCGTCTGGGCGGCGGAGCGCGGGGCGAGCCCACGTCCTCCTGAGGAAGCGGGGAGTGGAC   206
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  CCTGCGTCTGGGCGGCGGAGCGGCGGGGCGAGCCCACGTCCTCCTGAGGAAGCGGGGAGTGGAC  180

Qy    207  CATCGGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAATCGGTCTCTGGAGA      266
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  CATCGGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGA     240

Qy    267  TCACTGTAGAATTGTAGTGGATGAAAAATCAGGTCAGGTGACACTGGAAGATACCAGCAC    326
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  TCACTGTAGAATTGTAGTGGATGAAAAATCAGGTCAGGTGACACTGGAAGATACCAGCAC   300

Qy    327  CAGTGGAACAGTGATTGATTAACAACGTGAAGGTTGTTAAGAAGCAGAGCATGCCCTTTACAGAC  386
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  CAGTGGAACAGTGATCATCTTGGTGTACAGGAAGCAGAGCATGCCCTTTACAGAC           360

Qy    387  TGGGGATGTCATCTCACTTGGTGTACAGGAAGAATGAAACCGGAACAAGTGGCATACCT      446
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  TGGGGATGTCATCTCACTTGGTGTACAGGAAGAATGAACCGGAACACGTGGCATACCT       420

Qy    447  CTATGAATCTTTAAGTGAAAAGCAAGGCATGACACAAGATCCTTTGAAGCTAACAAGGA    506
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CTATGAATCTTTAAGTGAAAAGCAAGGCATGACACAAGATCCTTTGAAGATGGTGCCTTG   480

Qy    507  AAATGTGTTCCATGGGACCAAAGATACCTCAGGTGCAGGTGCAGGGGCGAGGGGCCGATCC   566
           |||||||||||||||||||||||||||||
Db    481  CTGTGTTGCCCAGCGGCTGGTCTAAAACTCCTGGGATCAAGT                      520

Qy    567  CCGGGTTCCCTCCGTCGTGCGCCCGCCACTCCAGGTGTCTTTGAGGAACCAGCCATCAAC   626
           ||||||||||||||||||||||||
Db    521  ---GATCCTCCCACCTTGGCCTTCCCCAAAGTATTGTGATTACA---                 559

Qy    627  ATCGACGTCAGACCTGTTCCCCACAGCGTCGGCCTTCTCCGGAGCCTTCCTCCTGCAGG    686
           ----------------------------------------------------------
Db    560  ----------------------------------------------------------     559

Qy    687  GCGAGAGCGTTCCTCCCAGTGTGGGGTCGTGGGGTCGTGGGTGGCATCTCCCCTAAAGGAAGTGG   746
           --------------------------------------------------
Db    560  -----------------GGGTCGGGGGTGGCATCTCCCCTAAAGGAAGTGG             597

Qy    747  TCCCTCCTGTGGCAAGTGATGAAGTTCCCAGCTTGCCTGCCTCTTCCCAGGACAGAAAGAC   806
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    598  TCCCCTCGTGTGGCAAGTGATGAAGTTCCCAGCTTGCCTGCCTCTTCCCAGGACAGAAAGAC    657

Qy    807  TGCGTCCTTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCCGGTGAAGAAGAA   866
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    658  TGCGTCCTTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCCGGAGGAAGAAGAA    717

Qy    867  AATGAGAGGAGGATGGGGACCTTGACCTGAACGGGCACGGCAGTTGTTGGTCGTGTTGGTCGTGCACACCGCGTAG  926
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    718  AATGAGAGGAGGATGGGGACCTTGACCTGAACGGGCAGTTGTTGGTCGTGTGTGTCGCACACCGCGTAG        777

Qy    927  AAATGCCCAAACCGTCCACGAGGACGTCAGAGCAGCGGCTGGGAGGCCAGACAAGATGGA   986
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    778  AAATGCCCAAACCGTCCACGAGGACGTCAGAGCAGCGGCTGGGAGGCCAGACAAGATGGA    837

Qy    987  GGAAGACGGTGACATGCAGCATCTGCCCAGGACCTGCGTGCGACTGCGTGAGTTGCAGCC  1046
           ||||||||||||| |||||||  |||||||||||||||| ||||||||||||||||||
Db    838  GGAGACGGTGACAGATGCACGATCATCATCTGCCCAGGACCTGCTGCGACTGCGTGAGTTGCAGCC  897

Qy   1047  CTGCATGCACACGTTCTCGCCGGCTTGCTACTCGGGCTGGATGGAGGCGCTCGTCCCTGTG  1106
           ||||||||||||||||| |||||||||||||||||||||||||||||| ||||||||||
Db    898  CTGCATGCACACGTTCTGCGGGCTTGCTACTCGGGCTGGATGGAGGCTCGTCCCTGTG      957

Qy   1107  TCCTACTTGCCGGTGTCCCGTGGAGCGGATCTGTAAAAACCACATCCTCAACAACCTCGT  1166
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    958  TCCTACTTGCCGGTGTCCCGTGGAGCGGATCTGTAAAAACCACATCCTCAACAACCTCGT   1017

Qy   1167  GGAAGCATACCTCATCAGACATCCAGACAAGAGTCGCAGTGAAGAGATGTGCAAAGTAT  1226
           ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1018  GGAAGCATACCTCATCCAGACATCCAGACAAGAGTCGCAGTGAAGAGATGTGCAAAGTAT   1077

Qy   1227  GGATGCCAGGAATAAAAATCACTCAAGACATGCTGCAGCCAAGTCAGGCGGTCTTTTTC  1286
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1078  GGATGCCCAGGAATAAAAATCACTCAAGACATGCTGCAGCCAAGTCAGGCGGTCTTTTTC   1137

Qy   1287  TGATGAAGAAGGAGTTCAGAGGACCTGCTGAGCTGTCGAGCTGTGACGTTGACAGTGAGTCCTC  1346
           |||||||||| ||||||||||||||||||||||||||||||||||| |||||||||||
Db   1138  TGATGAAGAAGGGAGTTCAGAGGACCTGCTGAGCTGTCGAGCTGTGACAGTGAGTCCTC      1197

Qy   1347  AGACATTAGCCAGCCATACGTCGTGTGCCCAGCGAGCCCCACTGCCCTGAGTCACAGGCGGC  1406
           |||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db   1198  AGACATTAGCCAGCCATACGTCGTGTGCCCAGCGAGCCCCAGAGGCAGGCGCGC           1257

Qy   1407  GCAGCCTCCCACTGCCAGCACGTCCGTCAGCGTCGAGACGCAGCAGTCCAGGATTACGTGTGCCCTCGCA  1466
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1258  GCAGCCTCCCACTGCCAGCACGTCCGTCAGCGTCGAGACGCAGCAGTCCAGGATTACGTGTGCCCTCTGCA  1317

Qy   1467  TGCACCTTCCAGTTCGCTGCTTGCTCTTCCAGCGACAGTGTGCGGTCGCGGAGCGGAGCG  1526
           |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   1318  TGCACCCCCAGTTCGCTGCTTGCTCTTCCAGCGACAGTGTGCGGTCGCGGAGCGGAGCG     1377

Qy   1527  AGGAAGCCACGCACCGCCCTGTGCTTCCAGCGACAGCCATGCCCGACCCGGAGCGCTTCTCGCCA  1586
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1378  AGGAAGCCACGCACCGCCCTGTGCTTCCAGCGACAGCCATGCCCGACCCGGAGCAGCGCCTTTCTGCCA  1437

Qy   1587  CGAGCAGGACCCGCCGGCTGTGCGGTCGTGTGCGGTCGTGCCTGCGCCCGTTTGTGAGCT  1646
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1438  CGAGCAGGACCCGCCGGCTGTGCGGTCGTGCCTGCGCCCGTTTGTGAGCT               1497

Qy   1647  CCTGTACTGGGGCTGCACCCGGCTGCACGGCGTGCCTGCGCCCGTTTGTGAGCT  1706
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1498  CCTGTACTGGGGCTGCACCCGGCTGCACGGCGTGCCTGCGCCCGTTTGTGAGCT         1557

Qy   1707  CAACCTGGGTGACAAGTGTCTGCAGCGCTGCTGAACAACAACAGCTACGAGTCAGATCAGACAT  1766
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1558  CAACCTGGGTGACAAGTGTCTGCAGCGCTGCTGAACAACAACAGCTACGAGTCAGATCAGACAT  1617

Qy   1767  CCTGAAGAATTACCTGCAACCAGAGGTTTGACATGGACCAGAGCCT  1826
           |||||||||||||||||||||||||||||||||||||||||||||
Db   1618  CCTGAAGAATTACCTGCAACCAGAGGTTTGACATGGACCAGAGCCT                   1677

Qy   1827  CGTGGCTCTGCCAGCGGGAGTGTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGT  1886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1678  CGTGGCTCTGCCAGCGGGAGTGTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGT     1737

Qy   1887  TCTGTGTTACTGCGTGGTGGCCTGCGCGAGCTTCGGTGAGCTGACCTATCAGTATCGCGACGAGA  1946
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1738  TCTGTGTTACTGCGTGGTGGCCTGCGTGGGCCTGCGCGAGCTTCGGTGAGCTGACCTATCAGTATCAGCAGAA  1797

Qy   1947  CATTCCTGCTTCCCAGTTGCCCAGTGGCCGTTGGTGCTGCACATCCCGTCCTGACTCGGGGCCG  2006
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1798  CATTCCTGCTTCCCAGTTGCCCAGTGGCCGTTGGTGCTGCACATCCCGTACTGGGGCCG        1857

Qy   2007  TAACTGCCGCACTCAGGTGAAAGTCACCACGCGCCATGAATTCAATCATATCGTGAACA  2066
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1858  ...
```

Db 778 AAATGCCAAACCGTCCAGAGGACGTCAGAGCAGCGGCTGGGAAGCCAGACAAGATGGA 837

Qy 987 GGAGACGCTGACATGCATCCTGCGCGGCTGCACGACTGCGTGAGTTTGCAGCC 1046
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 838 GGAGAGCGCTGACATGCATCCTGCGCTGCACGACTGCGTGAGTTTGCAGCC 897

Qy 1047 CTGCAATGCACACGTTCTGCGCGGCTACTCGGCGTGATGGAGCGCTCGTCGTCCCTGTG 1106
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 898 CTGCAATGCACACGTTCTGCGCGGCTACTCGGCGTGATGGAGCGCTCGTCGTCCCTGTG 957

Qy 1107 TCCTACCTGCGCGTGCTGTCCCGTGGAGCGGATCGTAAAAACCACATCCTCAACAACCTCGT 1166
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 958 TCCTACCTGCGCGTGCTGTCCCGTGGAGCGGATCGTAAAAACCACATCCTCAACAACCTCGT 1017

Qy 1167 GGAAGCATACCTCATCCAGCATCCAGACAACAAGAGTCGCAGTGAACAAGAGATGTGCAAAGTAT 1226
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1018 GGAAGCATACCTCATCCAGCATCCAGACAACAAGAGTCGCAGTGAACAAGAGATGTGCAAAGTAT 1077

Qy 1227 GGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTTCAGGGGTCTTTTTC 1286
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1078 GGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTTCAGGGGTCTTTTTC 1137

Qy 1287 TGATGAAGAGGGAGTTCAGAGGACCTGCTGAGCGTTGCACGTGACAGTGAGTCCTC 1346
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1138 TGATGAAGAGGGAGTTCAGAGGACCTGCTGAGCGTTGCACGTGACAGTGAGTCCTC 1197

Qy 1347 AGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAGGGCAGGCGGC 1406
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1198 AGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAGAGGCAGGCGGC 1257

Qy 1407 GCAGCCTTCCCCACTGCCCAGCACCCGGAGGGCGAGCGGCAGGAGCCCCCACAGCGCCCTGGGGGA 1466
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1258 GCAGCCTTCCCCACTGCCCAGCACCCGGAGGGCGAGCGGCAGGAGCCCCACAGCCCCTGGGGGA 1317

Qy 1467 TGCACCCTCCAGTCCGTCAGCGTGACGACAGCAGTCCAGGATTACGTGTCCCCTGCA 1526
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1318 TGCACCCCCCACGTCCGTCAGCGTGACGACAGCAGTCCAGGATTACGTGTCCCCTGCA 1377

Qy 1527 AGGAAGCCACCGCCCCTGTTGCACCTGCTTCCAGCCCATGCCGACGGCACCGGAGCGG 1586
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1378 AGGAAGCCACCGCCCCTGTTGCACCTGCTTCCAGCCCATGCCGACGGCACCGGAGCGG 1437

Qy 1587 CGAGCAGGACCCCGGGTGCGGCTGCCCCTCAGCAGTGTGCGGCCTTGCTGCCA 1646
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1438 CGAGCAGGACCCCGGGTGCGGCTGCCCCTCAGCAGTGTGCGGCCTTGCTGCCA 1497

Qy 1647 CCTGTACTGGGGCTGCACCCGGACCGGCTACGGCGTGCTGCTGGCCCCGGTTTTGTGAGCT 1706
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1498 CCTGTACTGGGGCTGCACCCGGACCGGCTACGGCGTGCTGCTGGCCCCGGTTTTGTGAGCT 1557

Qy 1707 CAACCTGGGTGACAAGTGTCTGGACGCGTGCTGCTGAACAACAGCAGTACGAGTCAGACAT 1766
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1558 CAACCTGGGTGACAAGTGTCTGGACGCGTGCTGCTGAACAACAGCAGTACGAGTCAGACAT 1617

Qy 1767 CCTGAAGAATTACCTGGCAACCAGAGGTTGACATGGAAAAACATGTTGACCGAGAGCCT 1826
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1618 CCTGAAGAATTACCTGGCAACCAGAGGTTGACATGGAAAAACATGTTGACCGAGAGCCT 1677

Qy 1827 CGTGGGCTCTCCAGCGCGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGCGGAGACACCGT 1886
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1678 CGTGGGCTCTCCAGCGCGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGCGGAGACACCGT 1737

Qy 1887 TCTGTGTTGACTGCTGGTGTGGCTCCGGCCAGTTCCGTGAGCTGACCTATCAGTATCGGCAGAGA 1946
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1738 TCTGTGTTGACTGCTGGTGTGGCTCCGGCCAGTTCCGTGAGCTGACCTATCAGTATCAGCAGAA 1797

Qy 1947 CATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCCGTGACTGCTACTCGGGGCCG 2006
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1798 CATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCCGTGACTGCTACTCGGGGCCG 1857

Qy 2007 TAACTGCCGCACTCAGTGAAAGCTGAAAGCTGCACCACGCCCATGAAATTCAATCATATCGTGAACA 2066
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1858 TAACTGCCGCACTCAGTGAAAGCTGAAAGCTCACCACGCCCATGAAATTCAATCATATCGTGAACA 1917

Qy 2067 GACAAGGTTCAAAAACTAAGCATCCAGAGCCCTGAGCAGCTTTCAGCACTGGAGGTGAA 2126
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1918 GACAAGGTTCAAAAACTAAGCATCCAGAGCCCTGAGCAGCTTTCAGCACTGGAGGTGAA 1977

Qy 2127 GAGAGCGTGTTTTTAAAATACAGAGACAAGCAGTCAAGGTGTTTTCACAGCCCCCTGAG 2186
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1978 GAGAGCGTGTTTTTAAAATACAGAGACAAGCAGTCAAGGTGTTTTCACAGCCCCCTGAG 2037

Qy 2187 GGAAGGACCAGGGTCTCCGACAGGTGCTCGGGGTGACTCTTCTGTGGAGCTTTTTAC 2246
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2038 GGAAGGACGCAGGGTCTCCGACAGGTGCTCGGGGTGACTCTTCTGTGGAGC--TTTAC 2095

Qy 2247 CCTCTGAGTGAGACCCTCCCCAGAGCCCGGGGGCGCAGCCCGGGGGCCCTCCTGGTGAGCGC 2306
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2096 CCTCTGAGTGAGACCCTCCCCAGAGCCCGGGGGCGCAGCCCGGGGGCCCTCCTGGTGAGCGC 2155

Qy 2307 TGGGCAGGGCTCGTGGTGCGCATCAGCAGCAGACGAGCCTTTCTGTAACATGCGGCCG 2366
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2156 TGGGCAGGGCTCGTGGTGCGCATCAGCAGCAGACGAGCCTTTCTGTAACATGCGGCCG 2215

Qy 2367 TCCCGCCGAGAGGGGCCAGGTTTTGCTTCTTTTGTACATTTTCCGAAACTACAGTTAAAGCAG 2426
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2216 TCCTGCCGAGAGGGGCCAGGTTTTGCTTCTTTTGTACATTTTCCGAAACTACAGTTAAAGCGG 2275

Qy 2427 AGTCTGTTTTCAGGAAAAGTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTC 2486
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2276 AGTCTGTTTTCAGGAAAAGTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTC 2335

Qy 2487 AGGAAGAGGGAGCATAAGTTACAGCCTACAGACGTACACAAATCCTGCTGTGGGAA 2546
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2336 AGGAAGAGGGAGCATAAGTTACAGCGTACACAGACGTACACAAATCCTGCTGTGGGAA 2395

Qy 2547 AACCACAGCATTTATCTATTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGA 2606
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2396 AACCACAGCATTTATCTATTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGA 2455

Qy 2607 TTTAAATGTCACAAACTGTAGCACAAATATATAATTATAATTTACAAATTGACTAAAA 2666
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2456 TTTAAATGTCACAAACTGTAGCACAAATATATAATTATAATTTACAAATTGACTAAAA 2515

RESULT 7
BD156548
LOCUS   BD156548   3138 bp   DNA   linear   PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD156548
VERSION   BD156548.1   GI:27862306
KEYWORDS   JP 2002191363-A/11391.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3138)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE   Primer for synthesizing full-length cDNA and use thereof
  JOURNAL   Patent: JP 2002191363-A 11391 09-JUL-2002;
            HELIX RESEARCH INSTITUTE
COMMENT   OS   Homo sapiens (human)
            PN   JP 2002191363-A/11391
            PD   09-JUL-2002
            PF   28-JUL-2000 JP 2000280990
            PI   TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
            PI   SAITO,
            PI   JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
            PI   KEIICHI NAGAI,TETSUJI OTSUKI
            PC   C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
            10,
            PC   C12N15/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
            Primer for synthesizing full-length cDNA and use thereof FH Key
            Location/Qualifiers

```
                    VQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPVVCRQCPE
                    YRRQAQPPHCPAPEGEPGAPQALGDAPPTSVSLTTAVQDYVCPLQGSHALCTCCFQP
                    MPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELNLGDKCLDGV
                    LNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLR
                    SFRELTYQYQQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQTRFKN"

ORIGIN

Query Match          84.4%;   Score 2262.4;   DB 6;   Length 3138;
Best Local Similarity 93.0%;   Pred. No. 0;
Matches 2454;   Conservative   0;   Mismatches   61;   Indels 125;   Gaps   3;

Qy    27  CTCTTGACAGCGGCGGCGGCGGCCAGCCGGTTCCGGGTTCGGGTTCGGGCGGCGGCGGGGATGTGAAT   86
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  CTCTTGACAGCGGCGGCGGCGGCCAGCCGGTTCCGGGTTCGGGTTCGGGCGGCGGCGGGGATGTGAAT   60

Qy    87  CCCGATGGAGCGGCCGAGGAAGCAAGCAGTCGCCGCGCCGCGCCCTGGGGACGGCT  146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  CCCGATGGAGCGGCCGAGGAAGCAAGCAGTCGCCGCGCCGCGCCCTGGGGACGGCT  120

Qy   147  CCTGGCGTCTGGGGCGGCGGAGGAGGGCGAGCCGCACGTCCTCCTGAGGAAGCGGGAGTGGAC  206
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CCTGGCGTCTGGGGCGGCGGAGGAGGGCGAGCCGCACGTCCTCCTGAGGAAGCGGGAGTGGAC  180

Qy   207  CATCGGGCCGGAGACGAGGTTGCGACCTTTCCTTCCCAGCAATAAACTGGTCTCTGGAGA  266
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  CATCGGGCCGGAGACGAGGTTGCGACCTTTCCTTCCCAGCAATAAACTGGTCTCTGGAGA  240

Qy   267  TCACTGTAGAATTGTAGTGGATGAAAAATCAGTCAGTGACTGACACTGAAGATACCAGCAC  326
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  TCACTGTAGAATTGTAGTGGATGAAAAATCAGTCAGTGACTGACACTGAAGATACCAGCAC  300

Qy   327  CAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGCAGACATGCCCTTTACAGAC  386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  CAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGCAGACATGCCCTTTACAGAC  360

Qy   387  TGGGGATGTCATCTACTTGTGTACAGGAAGAATGAACCGGAACAACAACGTGGCATACCT  446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  TGGGGATGTCATCTACTTGTGTACAGGAAGAATGAACCGGAACAACAACGTGGCATACCT  420

Qy   447  CTATGAATCTTTAAGTGAAAAGCAAGGCATGACACAACAAGAATCCTTTGAAGCTAACAAGGA  506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  CTATGAATCTTTAAGTGAAAAGCAAGGCATGACACAACAAGAATCCTTTGAAGATGTGCCTTG  480

Qy   507  AAATGTGTTCCATGGGACCAAAAGATACCTCAGTGCAGGGGCGCAGGGGCGAGGGGCCGATCC  566
          |||||||||||||||||||||||||||||||||       ||
Db   481  CTGTGTTGCCCAGGCTGGTCTAAAACTCCTGGGATCAGT----------------------  520

Qy   567  CCGGGTCCCTCCGTGTCGCCGCCACTCAGGTGTGCTTTGAGGAACCACAGCCATCAAC  626
          ||||||||||||||||      |||
Db   521  ---GATCCTCCACCTTGGCCTCCCAAAGTATTGTGATTACA-----------------  559

Qy   627  ATCGACGTCAGACCTCTTCCCCACAGCGTCTCGGCCTTCCTCCTCCCACGGAGCCTTCTCCTGCAGG  686
          |||||||||||||||||||||||||||||||||
Db   560  --------------------------------------------------------------  559

Qy   687  GCGAGAGCGTTCCTCCAGTGTGGGTCTCGGGGTGGTGGCATCTCCCCTAAAGGAAGTGG  746
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   560  --------------------GGGTCTCGGGGTGGCATCTCCCCTAAAGGAAGTGG  597

Qy   747  TCCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTTGCCTCAGCTTTGCCTCAGACAGAAAGAC  806
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   598  TCCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTTGCCTCAGCTTTGCCTCAGACAGAAAGAC  657

Qy   807  TGGCGTCCTTTCGTCGTTGGAACCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAA  866
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   658  TGGCGTCCTTTCGTCGTTGGAACCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAA  717

Qy   867  AATGAGGAGGATGGGACCTTGACCTGAACGGGCCAGTTGTCGTCGCACAACCGCGTAG  926
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   718  AATGAGGAGGATGGGACCTTGACCTGAACGGGCCAGTTGTTGGTCGCACAACCGCGTAG  777

Qy   927  AAATGCCCAAACCGTCCACGAGGACGTCAGAGCAGCAGCGGCTGGAAGCCAGACAAGATGGA  986
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  2005  TGAAGGAGAGCGGTGTTTTAAAATACAGAGACAAGGACGTGGTTTTCACAGCCCCC  2064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db  2183  TGAGGGAAGGGACCGCAGGGTCTCCGACAGGTGCTTCTGGGGTGACTCTTCTGTGGAGCTTT  2242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2065  TGAGGGAAGGGACCGCAGGGTCTCCGACAGGTGCTTCTGGGGTGACTCTTCTGTGGAGCTTT  2124

Qy  2243  TTACCCTCTGAGTGAGACCCTCCCCAGAGCCCCGGGGAGCCCGGAGCGGCCGCCCTCCTGGTGA  2302
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2125  TTACCCTCTGAGTGAGACCCTCCCCAGAGCCCCGGGGAGCCCGGAGCGGCCGCCCTCCTGGTGA  2184

Qy  2303  GCGCTGGGCAGGGCATCAGCAGCAGAGACGAAGCCTTTCTGTAACATGCG  2362
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2185  GCGCTGGGCAGGGCATCAGCAGCAGAGACGAAGCCTTTCTGTAACATGCG  2244

Qy  2363  GCCGTCCCGCCGAGAGGGGCAGTTTTGCTCTTTTGTACATTTTCCGAAACTACAGTTAAA  2422
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2245  GCCGTCCCGCCGAGAGGGGCAGTTTTGCTCTTTTGTACATTTTCCGAAACTACAGTTAAA  2304

Qy  2423  GCAGAAGTCTGTTTTCAGGAAAAGTTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATTG  2482
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2305  GCAGAAGTCTGTTTTCAGGAAAAGTTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATTG  2364

Qy  2483  TTTCAGGAGAAGGGGAGCATAAGTTTACAGCCTACAGGACGTACACATATCCTGCTGCTG  2542
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2365  TTTCAGGAGAAGGGGAGCATAAGTTTACAGCCTACAGGACGTACACATATCCTGCTGCTG  2424

Qy  2543  GGAAAACCACAGCATTTTATCTATTTTTATTTTTAATAGGTTTGGTGCTTATCTTCTAAT  2602
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2425  GGAAAACCACAGCATTTTATCTATTTTTATTTTTAATAGGTTTGGTGCTTATCTTCTAAT  2484

Qy  2603  AAGATTTAAATGTCACAAACTGTAGCACAAATAATAATTACAAATTGACA  2662
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2485  AAGATTTAAATGTCACAAACTGTAGCACAAATAATAATTATAAATTGACT  2544

Qy  2663  AAAA  2666
          ||||
Db  2545  AAAA  2548
```

```
RESULT 6
AX877222
LOCUS       AX877222                3138 bp    DNA       linear       PAT  17-DEC-2003
DEFINITION  Sequence 12127 from Patent EP1074617.
ACCESSION   AX877222
VERSION     AX877222.1   GI:40031958
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primers for synthesising full-length cDNA and their use
  JOURNAL   Patent: EP 1074617-A 12127 07-FEB-2001;
            Research Association for Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1..3138
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             65..1936
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAE89626.1"
                     /db_xref="GI:40031959"
                     /translation="MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRR
                     GCDLSFPSNKLVSGDHCRIVDEKSGQVLEDTSTSGTVINKLKVVKKQTCPLQTGDV
                     IYLVRKNEPEHNVAYLYESLSEKQGMTGDEVFEMVPCCVAQAGLKLLGSSDPPFLASQ
                     SIVITSGVGGGISPKGSGPSVASDEVSFASALPDKRTASFSSLLEPQDQDBLEPEKKK
                     MRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEETLTCIICODLLHDCVSL
                     QPCMHTFCAACYSGHWERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEED
```

complex, ER membrane component [Posttranslational
modification, protein turnover, chaperones]"
        /db_xref="CDD:COG5243"

ORIGIN

Query Match   93.2%; Score 2497.2; DB 9; Length 3189;
Best Local Similarity 98.5%; Pred. No. 0;
Matches 2545; Conservative 0; Mismatches 3; Indels 36; Gaps 1;

```
Qy    83 GAATCCCGATGGAGCGGCCGAGGAAGGCAAGCAGTCGCCGCCGCCCAGCCAGTCGGGGAC 142
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GAATCCCGATGGAGCGGCCGAGGAAGGCAAGCAGTCGCCGCCGCCCAGCCAGTCGGGGAC 60

Qy   143 GGCTCCTGCGTCTGGCGCGCGGAGGAGGCGAGCGCGCACGTCCTCCTGAGGAAGAGCGGGAGT 202
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GGCTCCTGCGTCTGGCGCGCGGAGGAGGCGAGCGCGCACGTCCTCCTGAGGAAGAGCGGGAGT 120

Qy   203 GGACCATCGGGCCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGGTCTCTG 262
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGACCATCGGGCCGGAGACCAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTCGGTCTCTG 180

Qy   263 GAGATCACTGTAGAATTGTAGTCGATGAAAAATCAGGTCAGGTGACACTGGAAGATACCA 322
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GAGATCACTGTAGAATTGTAGTCGATGAAAAATCAGGTCAGGTGACACTGGAAGATACCA 240

Qy   323 GCACCAGTGGAACAGTGATTAACAAGCTGAAGTTGTTAAGAGCAGACATGCCCTTTAC 382
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCACCAGTGGAACAGTGATTAACAAGCTGAAGTTGTTAAGAGCAGACATGCCCTTTAC 300

Qy   383 AGACTGGGGATGTCATCTACTTGGTGTACAGGAGAATGAACCGGACACAACAACGTGCAT 442
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGACTGGGGATGTCATCTACTTGGTGTACAGGAGAATGAACCGGACACAACAACGTGCAT 360

Qy   443 ACCTCTATGAATCTTTAAGTGAAAAAGCAAGGCATGACACAAGAATCCTTGAAGCTAACA 502
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ACCTCTATGAATCTTTAAGTGAAAAAGCAAGGCATGACACAAGAATCCTT---------- 410

Qy   503 AGGAAAATGTGTTCCATGGACCAAAGATACCTCAGGTGCAGGTGCAGGGGCGAGGGGCCG 562
                                 ||||||||||||||||||||||||||||||||||||||
Db   411 --------------------GATACCTCAGGTGCAGGTGCAGGGGCGAGGGGCCG 444

Qy   563 ATCCCCGGTCCCTCCGTCGTGCGCCGCCACTCAGGTGCGTTGAGGAACCACAGCCAT 622
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   445 ATCCCCGGTCCCTCCGTCGTGCGCCGCCACTCAGGTGCGTTGAGGAACCACAGCCAT 504

Qy   623 CAACATCGCAGTCAGACCTCTTCCCCACAGCCTCGGCCTCTTCCACGGAGCCTTCTCCTG 682
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   505 CAACATCGCAGTCAGACCTCTTCCCCACAGCCTCGGCCTCTTCCACGGAGCCTTCTCCTG 564

Qy   683 CAGGGCGAGAGCCGTTCCTTCCAGTTGTGGGTCTGGGGGTCTGGGGGTGTGGCATCTCCCCTAAAGGAA 742
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   565 CAGGGCGAGAGCCGTTCCTTCCAGTTGTGGGTCTGGGGGTGTGGCATCTCCCCTAAAGGAA 624

Qy   743 GTGGTCCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTGCCTCAGCTTCCCAGCTTGCCTCAGCTTCCCAGCTTGCCTCAGACAGAA 802
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   625 GTGGTCCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTGCCTCAGCTTCCCAGCTTGCCTCAGACAGAA 684

Qy   803 AGACTGGCTCCTTTCGTCGTTGGAACCCAGGATCAGGAGATTTGGAGCCCGTGAAGA 862
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   685 AGACTGGCTCCTTTCGTCGTTGGAACCCAGGATCAGGAGATTTGGAGCCCGTGAAGA 744

Qy   863 AGAAAATGAGAGAGATGGGAACCTTGACCTGGTTGTTGGTCGCACACCGC 922
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   745 AGAAAATGAGAGAGATGGGAACCTTGACCTGGTTGTTGGTCGCACACCGC 804

Qy   923 GTAGAAATGCCCAAACCGTCCAAGAGCAGTCAGAGCAGCGGCTCAGAGGAAGCCAGACAGA 982
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   805 GTAGAAATGCCCAAACCGTCCAAGAGCAGTCAGAGCAGCGGCTCACCAGGCCAGACAGA 864

Qy   983 TGGAGGACGCGTCGACATGTCATCATCTGCCAGGACCTGCTCACGACCTGCGTGAGTTTGC 1042
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   865 TGGAGGACGCGTCGACATGTCATCATCTGCCAGGACCTGCTCACGACCTGCGTGAGTTTGC 924
```

```
Qy  1043 AGCCCTGCATGCACACGTTCTGCGCGGCGTTGCTACTCGGCTCGGATGGAGCGGCTCGTCCC 1102
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   925 AGCCCTGCATGCACACGTTCTGCGCGGCGTTGCTACTCGGCTCGGATGGAGCGGCTCGTCCC 984

Qy  1103 TGTGTCCTACCTGCCCGTGTCCGTGTCCGTGGAGCCGATCTGTGAGCGGATCTGTAAAAACCATCCTCAACAACC 1162
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   985 TGTGTCCTACCTGCCCGTGTCCGTGTCCGTGGAGCCGATCTGTGAGCGGATCTGTAAAAACCACATCCTCAACAACC 1044

Qy  1163 TCGTGGAAGCATACTTCATCCAGCATCCAGACAAGAGTCGCAGTGAAGAAGATGTGCAAA 1222
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1045 TCGTGGAAGCATACTTCATCCAGCATCCAGACAAGAGTCGCAGTGAAGAAGATGTGCAAA 1104

Qy  1223 GTATGGATGCCAGGAATAAAATCACTCAAGACATGTCGCAGGCGGCGGGTCTT 1282
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1105 GTATGGATGCCAGGAATAAAATCACTCAAGACATGTCGCAGGCGGGTCTT 1164

Qy  1283 TTTCTGATGAAGAAGGGAGTTCAGAGGACCTGCTCTGGGAGCGTGTCAGTGAGT 1342
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1165 TTTCTGATGAAGAAGGGAGTTCAGACCTGCTCTGGGAGCGTGTCAGTGAGT 1224

Qy  1343 CCTCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTCACGAAGCCAGG 1402
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1225 CCTCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTCACGAAGCCAGG 1284

Qy  1403 CGGCGCAGCCTCCCCACTGCCCAGCACCCGGAGCCGAGCCAGGCCCACGG 1462
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1285 CGGCGCAGCCTCCCCACTGCCCAGCACCCGGAGCCGAGCCAGGCCCACTGG 1344

Qy  1463 GGGATGCACCCTCCAGTCCGTCAGCCAGCAGCAGTCCAGATTACGTGTGCCCTC 1522
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1345 GGGATGCACCCTCCAGTCCGTCAGCCAGCAGCAGTCCAGATTACGTGTGCCCTC 1404

Qy  1523 TGCAAGGAAGCCAGCCACGCCCTGTTGCACCTGCTTCCAGCAGTGTGCGCGTCTGCCTGCAGCCTTTCTT 1582
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1405 TGCAAGGAAGCCAGCCACGCCCTGTTGCACCTGCTTCCAGCAGTGTGCGCGTCTGCCTGCAGCCTTTCT 1464

Qy  1583 AGCGCGAGCAGGACCCCGGCGGTGCAGGTGCAGGTGCAGCGGAGAGTGG 1642
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1465 AGCGCGAGCAGGACCCCGGCGGTGCAGGTGCAGCGGAGAGTGG 1524

Qy  1643 GCCACTGTACTGGGGCTGCACACCGGCTGCTACGGCTGCCTGGCCCCGTTTGTG 1702
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1525 GCCACTGTACTGGGGCTGCACGGCTGCTACGGCTGCCTGGCCCCGTTTGTG 1584

Qy  1703 AGCTCAACCTGGGTGACAAGTGTCTGACGGCGTGCTGGCAACACAGCAGTCGAGTCAG 1762
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1585 AGCTCAACCTGGGTGACAAGTGTCTGACGGCGTGCTGGCAACACAGCAGTCGAGTCAG 1644

Qy  1763 ACATCCTGAAGAATTACCTGCCAACCAGAGGTTTGACATGGAAAAACATGTTGACCGAGA 1822
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1645 ACATCCTGAAGAATTACCTGCAACCAGAGGTTTGACATGGAAAAACATGTTGACCGAGA 1704

Qy  1823 GCCTCGTGTGCCTCTCCAGCGGGGAGTGTTTCTGCTGTCGTTGATTACAGAGTCACGGAGACA 1882
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1705 GCCTCGTGTGCCTCTCCAGCGGAGTGTTTCTGCTGTCGTTGATTACAGAGTCACGGAGACA 1764

Qy  1883 CCGTTCGTGTTACTGCTGTGTGCCTCGCAGCTTCCGTGACCTTCAGCTTTCAGCTGTGAAGA 1942
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1765 CCGTTCGTGTTACTGCTGTGTGCCTCGCAGCTTCCGTGACCTTCAGCTTTCAGCTGTGAAGC 1824

Qy  1943 AGAAACATTCCTGCTTCCGTAACATCCCGTGCTACTGGG 2002
         ||||||||||||||||||||||||||||||||||||||
Db  1825 AGAGACATTCCTGCTTCCGTAACATCCCGTGCTACTGGG 1884

Qy  2003 GCCGTAACTGCGCGCAAACCGTCCACCACCGCCATGAAAATTCAATCATATCTGTG 2062
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1885 GCCGTAACTGCGCGCACTCAGGTGAAAGCTCACCACCGCCATGAAAATTCAATCATATCTGTG 1944

Qy  2063 AACAGACAAGGTTCAAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAGG 2122
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1945 AACAGACAAGGTTCAAAAACTAAGCAGCCCTGAGCAGCTTTCAGCACTGGAGG 2004

Qy  2123 TGAAGAGAGCCGTGTTTTTAAAATACAGAGACAAGACGTCAAGGTGTTTCACAGCCCCC 2182
```

```
Db  2069  TGGAGGTGAAGAGAGCGTGTGTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTTCACA  2128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2177  GCCCCCTGAGGGAGGGACGCAGGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCTGTGG   2236
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2129  GCCCCCTGAGGGAGGGACGCAGGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCTGTGG   2188

Qy  2237  AGCTTTTACCCTCTGAGTGAGACCCTCCCAGAGCCCGGGGGCCGAGCCCGCCCTCC     2296
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2189  AGCTTTTTACCCTCTGAGTGAGACCCTCCCAGAGCCCGGGGGCCGAGCCCGCCCTCC    2248

Qy  2297  TGGTGAGCGCTGGGCAGGGCTCGTGGTGGCATCAGCAGACGAAGCCTTCTGTAA      2356
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2249  TGGTGAGCGCTGGGCAGGGCTCGTGGTGGCATCAGCAGACGAAGCCTTCTGTAA     2308

Qy  2357  CATGCGGCCGTCCCGCCGAGGGGCCAGTTTGCTCCTTTTGTACATTTCCGAAACTACA   2416
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2309  CATGCGGCCGTCCCGCCGAGGGGCCAGTTTGCTCCTTTTGTACATTTCCGAAACTACA  2368

Qy  2417  GTTAAAGCAGAAGTCTGTTTTCAGGAGGAAGTTCAAGGGGCAAGTTTATCAAAA      2476
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2369  GTTAAAGCAGAAGTCTGTTTTCAGGAAAAGTTCAAGGGGCAAGTTTATCAAAA     2428

Qy  2477  ACATTGTTTCAGGAGAGGAGCATAAGTTTACAGCTCAGGACGTACACAATATCCTG    2536
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2429  ACATTGTTTCAGGAGAGGAGCATAAGTTTACAGCCTACAGGACGTACACAATATCCTG 2488

Qy  2537  CTGCTGGGAAAACCACAGCATTTTATCTATTTTTATTTTTAATAGGTTTGGTGCTTATCT 2596
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2489  CTGCTGGGAAAACCACAGCATTTTATCTATTTTTATTTTTAATAGGTTTGGTGCTTATCT 2548

Qy  2597  TCTATTAAGATTTAAATGTCCAAAACTGTAGCACAAATAATATATTATAATTTACAAA   2656
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2549  TCTATTAAGATTTAAATGTCACAAACTGTAGCACAAATAATATATTATAATTTACAAA  2608

Qy  2657  TTGACAAAAA  2666
          |||| ||||
Db  2609  TTGACTAAAA  2618
```

RESULT 5
BC012072

LOCUS   BC012072   3189 bp   mRNA   linear   PRI 04-OCT-2003
DEFINITION Homo sapiens checkpoint with forkhead and ring finger domains, mRNA
            (cDNA clone MGC:19963 IMAGE:4650348), complete cds.
ACCESSION   BC012072
VERSION   BC012072.1 GI:15082329
KEYWORDS   MGC.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 3189)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE   22388257
    PUBMED   12477932
REFERENCE   2 (bases 1 to 3189)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 29 Row: n Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 8922674.
FEATURES   Location/Qualifiers
     source   1..3189
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:19963 IMAGE:4650348"
                     /tissue_type="Placenta, choriocarcinoma"
                     /clone_lib="NIH_MGC_21"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene   1..3189
                     /gene="CHFR"
                     /note="synonyms: FLJ10796, RNF116"
                     /db_xref="LocusID:55743"
                     /db_xref="MIM:605209"
     CDS   9..1967
                     /codon_start=1
                     /product="CHFR protein"
                     /protein_id="AAH12072.1"
                     /db_xref="GI:15082330"
                     /db_xref="LocusID:55743"
                     /translation="MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRR
                     GCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTGSGTVINKLKVKKQTCPLQTGDV
                     IYLVRKNEPEHNVAYLYESLSEKQGMTQESFDTSGAGARGADRPVPPSSPATQVCF
                     EEPQPSTSDLFPTASASSTEPSPAGRERSSCCSGGGGISPKGSGPSVASDEVSSF
                     ASALPDRKTASFSSLEPQDEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDV
                     RAAAGKPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV
                     ERICKNHILNNLVEAYLIQHPDKSRSEEDVSMDARNKITQDMLQPKVRRSFSDEGS
                     SEDLLSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPS
                     TSVSLTTAVQDVYCPLQGSHALCTCCFQPMPDRRVEREQDPRVAPQQCAVCLQPFCHL
                     YWGCTRGCYGCLAPFCELNLGDHALCICDGVLNNNSYESDILKNYLATRGLJTWKNMLTES
                     LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCY
                     WGRNCRTRFKN"
     misc_feature   120..317
                     /note="FHA; Region: FHA domain. The FHA
                     (Forkhead-associated) domain is a phosphopeptide binding
                     motif"
                     /db_xref="CDD:pfam00498"
     misc_feature   879..1010
                     /note="HRD1; Region: COG5243, HRD1, HRD ubiquitin ligase
```

us-10-048-046-1.rge

```
Query Match            95.8%;    Score 2566.4;   DB 9;   Length 2639;
Best Local Similarity  98.6%;
Matches 2613;  Conservative 0;  Mismatches 0;  Pred. No. 0;  Indels 36;  Gaps 1;

Qy    17 GCCGCAATGTCTTCTTGACACAGCGGCGGCGGCCGAGCCGGTTCCGGGGTTCGGCGCGGCGGGGCGG    76
         ||||||||||||| |||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db     5 GCCGCAATGTCTCTTGACAGCAGCGGCGGCGGCCGAGCCGGTTCCGGGGTTCGGCGCGGCGGGGCGG    64

Qy    77 GGATGTGAATCCGATGGAGCGGCGGATGGAGCCGAGGAAGCCAAGCAGAGCAGTCGCCGCCGCAGCCCT   136
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    65 GGATGTGAATCCGATGGAGCGGCGGATGGAGCCGAGGAAGCCAAGCAGAGCAGTCGCCGCCGCAGCCCT   124

Qy   137 GGGGACGGCTCCTGCGTCTGGGGCGGCCGGAGGAGGGCCGAGCCCGACACGTGTCCTCCTGAGGAAGC   196
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   125 GGGGACGGCTCCTGCGTCTGGGGCGGCCGGAGGAGGGCCGAGCCCGACACGTGTCCTCCTGAGGAAGC   184

Qy   197 GGGAGTGGACCATCGGCGCGGAGACGAGGTTGCGACCTTTCCTTCCCAGCAATAAACTGG   256
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   185 GGGAGTCGGACCATCGGCGCGGAGACGAGGTTGCGACCTTTCCTTCCCAGCAATAAACTGG   244

Qy   257 TCTCTGGAGATCACTGTAGAATTGTAGTGATGAAAATCAGGTCAGGTGACACTGGAAG   316
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   245 TCTCTGAGATCACTGTAGAATTGTAGTGATGAAAATCAGGTCAGGTGACACTGGAAG   304

Qy   317 ATACCAGCACCAGTGAAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGACAATGCC   376
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   305 ATACCAGCACCAGTGAAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGACAATGCC   364

Qy   377 CTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAGGAAGAATGAACCGAACAACG   436
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   365 CTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAGGAAGAATGAACCGAACAACG   424

Qy   437 TGGCATACCTCTATGAATCTTAAGTGAAAAGCAAGCATGACACAAGATCCTTTGAAG   496
         |||||||||||||||||||||||||||||||||||||||||||
Db   425 TGGCATACCTCTATGAATCTTAAGTGAAAAGCAAGCATGACACAAGATCCTTT----   480

Qy   497 CTAACAAGGAAAATGTGTTCCATGGGACCAAGATACCTCAGGTGCAGGTGCAGGGCGAG   556
                                  |||||||||||||||||||||||||||||||||
Db   481 ------------------GATACCTCAGGTGCAGGTGCAGGGCGAG   508

Qy   557 GGGCCGATCCCGGGGTCCCTTCGTCCCGCCCTCGCCCACTCAGGTGTGCTTTGAGGAACCAC   616
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   509 GGGCCGATCCCGGGGTCCCTTCGTCCCGCCCTCGCCCACTCAGGTGTGCTTTGAGGAACCAC   568

Qy   617 AGCCATCAACATGACACGTCAGACCTTCTTCCCCACAGCCTCGGCCTCTTCCACGGAGCCTT   676
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   569 AGCCATCAACATGACACGTCAGACCTTCTTCCCCACAGCCTCGGCCTCTTCCACGGAGCCTT   628

Qy   677 CTCCTGCAGGGCGAGAGCGTTCCTCCAGTTGTGGGTCTGGGGGTCGGGGTGGTGGCATCTCCCCTA   736
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   629 CTCCTGCAGGGCGAGAGCGTTCCTCCAGTTGTGGGTCTGGGGGTCGGGGTGGTGGCATCTCCCCTA   688

Qy   737 AAGGAGGTGGTCCCCTGTCGTGGCAAGTGATGAAGTCTCCAGCTTTGCCTCAGCTCTCTCCCAG   796
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   689 AAGGAGGTGGTCCCCTGTCGTGGCAAGTGATGAAGTCTCCAGCTTTGCCTCAGCTCTCTCCCAG   748

Qy   797 ACAGACACCGTTCGTCCTTTTCGTCGTGTTGGAACCCCAGATCAGGAGGATTTGGAGCCCG   856
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   749 ACAGAACACCGTTCGTCCTTTTCGTCGTGTTGGAACCCCAGATCAGGAGGATTTGGAGCCCG   808

Qy   857 TGAAGGAGAAAATGAGGAGGATGGGGACCTTGACCTGACCTGAGCTCAGACGCGCAGTTGGTCGCAC   916
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   809 TGAAGGAGAAAATGAGGAGGATGGGGACCTTGACCTGACCTGAGCTCAGACGCGCAGTTGGTCGCAC   868

Qy   917 AACCCGCTAGAAATGCCCAAACCGTCCAAACCGTCCAGAGCGGCTCGGGAGGCCAG   976
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   869 AACCCGCTAGAAATGCCCAAACCGTCCAGACGGCTCAGAGCGGCTCGGGAGGCCAG   928

Qy   977 ACAAGATGGAGGAGACGCTGACATGCATCATCCGCCAGGAACTCAGCTGCTGCACGACTGCGTGA  1036
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   929 ACAAGATGGAGGAGACGCTGACATGCATCATCCGCCAGGAACTCAGCTGCTGCACGACTGCGTGA   988


Qy  1037 GTTTGCAGCCCTGCATGCACACGTTCTGGGCGGCTTGCTACTCGGCTGGCTGCTACTCGGCTGGCTGGGATGGAGCGCT  1096
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   989 GTTTGCAGCCCTGCATGCACACGTTCTGCTACTCGGCTGGCTGCTACTCGGCTGGCTGGGATGGAGCGCT  1048

Qy  1097 CGTCCCTGTGTCCTACCTGCGCTGCCTGTCCCGTGAGCGGGATCTGTAAAACCACATCCTCA  1156
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1049 CGTCCCTGTGTCCTACCTGCGCTGCCTGTCCCGTGGAGCGGGATCTGTAAAAACCACATCCTCA  1108

Qy  1157 ACAACCTCGTGGAAGCATACCTCATCCAGCATCCATCCAGACAAGAAGAGTCGCAGTGAAGAAGATG  1216
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1109 ACAACCTCGTGGAAGCATACCTCATCCAGCATCCATCCAGACAAGAACAAGAGTCGCAGTGAAGAAGATG  1168

Qy  1217 TGCAAAGTATGGATGGCCAGGAATAAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGC  1276
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1169 TGCAAAGTATGGATGGCCAGGAATAAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGC  1228

Qy  1277 GGTCTTTTTCTGATGAGAGGAGTTCAGAGACCTGCTGGAGCGTGTCAGACGTTGACA  1336
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1229 GGTCTTTTTCTGATGAGAGGAGTTCAGAGGACCTGCTGGAGCGTGTCAGACGTTGACA  1288

Qy  1337 GTGAGTCCTCAGACATTAGCAGCCATACGTCGTGTGCCGGCAGTGTCCGGCAGTGTCCGGTACGAA  1396
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1289 GTGAGTCCTCAGACATTAGCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACGAA  1348

Qy  1397 GGCAGCGGCGGCCAGCCTCCCCACTGCCCAGGAGCGGGAGGGCGGAGCCCCACAGG  1456
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1349 GGCAGCGGCGGCCAGCCTCCCCACTGCCCAGGAGCGGGAGGGCGGAGCCCCACAGG  1408

Qy  1457 CCCTGGGGGATGCACCCTCCCAGCCTCCGACGACAGCAGCTCAGGATTACGTGT  1516
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1409 CCCTGGGGGATGCACCCTCCCAGCCTCGTCAGCGTGACGACAGCAGTCCCGGCAGTTACGTGT  1468

Qy  1517 GCCCTCTGCAAGGAGGCCACGGCCGCCGTGTCGCCCCCAGTGTCGCGGTCTCGCCTCGCAGC  1576
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1469 GCCCTCTGCAAGGAGGCCACGGCCGCCGTGTCGCTGCCCCCAGTGTCGCGGTCTCGCCTCGCAGC  1528

Qy  1577 GAGCGGAGCCGAGCGAGCAGGAGGACCCGGTGTCGCCGGTGTCGCGGTCTCGCCTCGCAGC  1636
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1529 GAGCGGAGCCGAGCGAGCAGGAGGACCCGGTGTCGCCGGTGTCGCGGTCTCGCCTCGCAGC  1588

Qy  1637 CTTTCTGCACCTGTACTGGGGCTTGCACCCGGGCTGCTACGGCTGCTGGCGCCCCGT  1696
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1589 CTTTCTGCACCTGTACTGGGGCTTGCACCCGGGCTGCTACGGCTGCTGGCGCCCCGT  1648

Qy  1697 TTTGTGAGCTCAACTGGGTGACAAGTGTCTGGACAGTGTCTGGACGGTGCTGAACACAACAGCTACG  1756
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1649 TTTGTGAGCTCAACTGGGTGACAAGTGTCTGGACGGTGCTGAACACAACAGCTACG  1708

Qy  1757 AGTCAGACATCCTGAAGAATTACCTGGCAACCAGAGTTTGACATGGAAAAAACATGTTGA  1816
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1709 AGTCAGACATCCTGAAGAATTACCTGGCAACCAGAGTTTGACATGGAAAAAACATGTTGA  1768

Qy  1817 CCGAGAGCCTCGTGGCTCTCCAGCGGGGGAGAGTGTTTCTCGTGTCTGATTACAGAGTCACGG  1876
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1769 CCGAGAGCCTCGTGGCTCTCCAGCGGGGGAGAGTGTTTCTCGTGTCTGATTACAGAGTCACGG  1828

Qy  1877 GAGACACCGTTCTGTGTTACTGCGTTGGCCTGTGGCCGTAACATCCCGTCCGACTCAGT  1936
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1829 GAGACACCGTTCTGTGTTACTGCGTTGGCCTGTGGCCGTAACATCCCGTCCGACTCAGT  1888

Qy  1937 ATCGGCAGAACATTCCTGCTTCCGAGTTGCCGAGTTGCCGTAACATCCCGTCCGACTGCT  1996
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1889 ATCGGCAGAACATTCCTGCTTCCGAGTTGCCGAGTTGCCGTAACATCCCGTAACATCCCGTCCGACTGCT  1948

Qy  1997 ACTGGGCCGTAACTGCCCGACTCAGGTGAAAGTCACCACGGCCATGAAATTCAATCATA  2056
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1949 ACTGGGCCGTAACTGCCCGACTCAGGTGAAAGTCACCACGGCCATGAAATTCAATCATA  2008

Qy  2057 TCTGTGAACAGACAAGGTTCAAAAACTAAGACATCCAGAGGCCCTGACGACTTTCAGCAC  2116
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2009 TCTGTGAACAGACAAGGTTCAAAAACTAAGACATCCAGAGGCCCTGACGACTTTCAGCAC  2068

Qy  2117 TGGAGGTGAAGAAGACCGTGTTTTAAAATACAGAGAACAAGACAGCACGTCAAGGTGTTTTCACA  2176
```

Db 2489 CTGCTGGAAACCACCAGCAGCATTTTACCTATTTTTATTTTTAATAGGTTTGGTGCTTATCT 2548
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2597 TCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATAATTATAATTTATAATTTACAAA 2656

Db 2549 TCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATAATTATAATTTATAATTTACAAA 2608

Qy 2657 TTGACAAAAA 2666
         ||||| |||
Db 2609 TTGACTAAAA 2618

RESULT 4
AK027687
LOCUS   AK027687   2639 bp   mRNA   linear   PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ14781 fis, clone NT2RP4000455, weakly similar
            to TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP0.
ACCESSION   AK027687
VERSION   AK027687.1 GI:14042552
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,
            Matsunawa,H., Ishii,S., Kawai,Y., Saito,Y., Yamamoto,J.,
            Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2639)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing;
            Research Association for Biotechnology; cDNA library construction,
            5'- & 3'-end one pass sequencing and clone selection: Helix
            Research Institute (supported by Japan Key Technology Center etc.)
            and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES   Location/Qualifiers
     source   1..2639
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RP4000455"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2RP4"
                     /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
     CDS   79..2037
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB55297.1"
                     /db_xref="GI:14042553"
                     /translation="MERPEEGKQSPPPQPWGLLRLGAEEGSPHVLKRREWTIGRRR
                     GCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINLKVVKQTCPLQTGDV
                     IVLVYRKNEPEHNVAYLYESLSEKQGMTQESFDTSGAGRGADPRVPPSSPATQVCF
                     EERQPSTSTSDLFPTAGASSTEPSPAGRERSSCSGGGISPKGSGSPSVASDEVSSF
                     ASALPDRKTASFSSLEFPDQEDLFNVKKMRGDGDLDLNGQLIVAQPRRNAQTVHEDV
                     RAAAGKPDKMEFTLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV
                     ERICKNHILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLPKVRRSFSDEEGS
                     SEDLLELSDVDSBSSDISQPYVVRCQCPEYRRQAAQPPHCEPAPEGEPGAPQALGDAPS
                     TSVSLTTAVQDYVCPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHL
                     YWGCTRTGCYGLCLAPFCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTBS
                     LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELITYQVRQNIPASELPVAVTSRPDCY
                     WGRNCRTQVKAHHAMKFNHICSQTRFKN"
ORIGIN

Qy 1409 [alignment]   1468
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   CCCTGGGGATGCAGCTCCGTCAGCTGCTCAGCCTGACGCAGCAGTCCAGGATTACGTGT 1468

Qy 1517 GCCCTCTGCAAGGAAGCCACGCCCACGCCTGCTCTGCACCTCGCCTGCTCCCGACACCGGA 1576
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1469 GCCCTCTGCAAGGAAGCCACGCCCACGCCTGCTCTGCACCTCGCCTGCTCCCGACACCGGA 1528

Qy 1577 GAGCGGAGCGGAGCGAGCAGCAGCAGCAGCCGACCCGCGTGTCGGTGTCGGGTCTGCCTGCCGCAGC 1636
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1529 GAGCGGAGCGGAGCGAGCAGCAGCAGCAGCCGACCCGCGTGTCGGTGTCGGGTCTGCCTGCCGCAGC 1588

Qy 1637 CTTTCTGCCACCTGTACTGGGGCTGCACCCGGCTGCTACGGCTGCTGCCTGGCCCCGT 1696
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1589 CTTTCTGCCACCTGTACTGGGGCTGCACCCGGCTGCTACGGCTGCTGCCTGGCCCCGT 1648

Qy 1697 TTTGTGAGCTCAACCTGGGTGACAGTGTCTGGACGGCCGTGCTGAACAACAGCTACG 1756
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1649 TTTGTGAGCTCAACCTGGGTGACAGTGTCTGGACGGCCGTGCTGAACAACAGCTACG 1708

Qy 1757 AGTCAGACATCCTGAAGAATTACCTGGCAACCAGAGGTTTGACATGGAAAAACATGTTGA 1816
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1709 AGTCAGACATCCTGAAGAATTACCTGGCAACCAGAGTTTGACATGGAAAAACATGTTGA 1768

Qy 1817 CCGAGAGCCTCGTGGCTCTCCAGCGGGGAGTGTTTCTGCTCTGATTACAGAGTCACGG 1876
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1769 CCGAGAGCCTCGTGGCTCTCCAGCGGGGAGTGTTTCTGCTCTGATTACAGAGTCACGG 1828

Qy 1877 GAGACACCGTTCTGTGTTACTGCTGCTGGCCTGGCCGCAGTTCCGTGAGCTGACCTATCAGT 1936
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1829 GAGACACCGTTCTGTGTTACTGCTGCTGGCCTGGCCGCAGTTCCGTGAGCTGACCTATCAGT 1888

Qy 1937 ATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCTGACTGCT 1996
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1889 ATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCTGACTGCT 1948

Qy 1997 ACTCGGGCCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATA 2056
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1949 ACTCGGGCCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATA 2008

Qy 2057 TCTGTGAACAGACAAGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCAC 2116
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2009 TCTGTGAACAGACAAGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCAC 2068

Qy 2117 TGGAGGTGAAGAGAGAGAGCTGTGTTTTAAAATACAGAGACAACAAGCACGTCAAGGTGTTTCACA 2176
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2069 TGGAGGTGAAGAGAGAGAGCTGTGTTTTAAAATACAGAGACAACAAGCACGTCAAGGTGTTTCACA 2128

Qy 2177 GCCCCCTGAGGGAAGGACGCAGGGGTCTCCGACAGGTGACTCTTCGTGG 2236
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2129 GCCCCCTGAGGGAAGGACGCAGGGGTCTCCGACAGGTGACTCTTCGTGG 2188

Qy 2237 AGCTTTTTACCCTCTGAGTGAGACCCTCCCAGAGCCCCAGCCCCGCCCTCC 2296
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2189 AGCTTTTTACCCTCTGAGTGAGACCCTCCCAGAGCCCCAGCCCCGCCCTCC 2248

Qy 2297 TGGTGAGCGGCTGGGCAGGGGCTCGTGGTGGCATCAGCAGCAGAAGCCTTTCGTAA 2356
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2249 TGGTGAGCGGCTGGGCAGGGGCTCGTGGTGGCATCAGCAGCAGAAGCCTTTCGTAA 2308

Qy 2357 CATGCCGCCGTCCCGAGAGGGGCAGTTTGCTCTTTTGTACATTTTCCGAAACTACA 2416
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2309 CATGCCGCCGTCCCGAGAGGGGCAGTTTGCTCTTTTGTACATTTTCCGAAACTACA 2368

Qy 2417 GTTAAAGCAGAAGTCTGTTTTCAGGAAAAGTTCAAGGAGAAGGGCAAGTTTATCAAAA 2476
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2369 GTTAAAGCAGAAGTCTGTTTTCAGGAAAAGTTCAAGGAGAAGGGCAAGTTTATCAAAA 2428

Qy 2477 ACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGGCTACACAATATCCTG 2536
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2429 ACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGGCTACACAATATCCTG 2488

Qy 2537 CTGCTGGAAAACCACCAGCAGCATTTTATCTATTTTTATTTTTAATAGGTGCTTATCT 2596

```
Qy   2657 TTGACAAAAA 2666
          |||||
Db   2609 TTGACTAAAA 2618


RESULT 3
BD156534
LOCUS       BD156534                2639 bp    DNA     linear   PAT 17-JAN-2003
DEFINITION  Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD156534
VERSION     BD156534.1  GI:27862292
KEYWORDS    JP 2002191363-A/11377.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2639)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primer for synthesizing full-length cDNA and use thereof
  JOURNAL   Patent: JP 2002191363-A 11377 09-JUL-2002;
            HELIX RESEARCH INSTITUTE
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002191363-A/11377
            PD   09-JUL-2002
            PF   28-JUL-2000 JP 2000280990
            PI   TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
            PI   SAITO,
            PI   JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
            PI   KEIICHI NAGAI,TETSUJI OTSUKI
            PC   C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N5/15,C12N5/ PC
            10,
            PC   C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
            Primer for synthesizing full-length cDNA and use thereof FH    Key
            Location/Qualifiers
FT   CDS          (79)..(2034).
FEATURES             Location/Qualifiers
     source          1..2639
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN
Query Match     95.8%;  Score 2566.4;  DB 6;  Pred. No. 0;
Best Local Similarity    98.6%;
Matches 2613;  Conservative     0;  Mismatches     1;  Indels   36;  Gaps     1;
```

```
Qy    377 CTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGAAGAGAATGAACCGGAACACACG 436
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    365 CTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGAAGAGAATGAACCGGAACACACG 424

Qy    437 TGGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGGCATGACACAACAAGAATCCTTGAAG---- 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    425 TGGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGGCATGACACAACAAGAATCTTT 480

Qy    497 CTAACAAGGAAAATGTGTTCCATGGGACCAAGATAACTCAGTGCAGTGCAGGGCGAG 556
          |||||||                 ||||||||||||||||||||||||||||||||
Db    481 --------------------GATACCTCAGGTGCAGTGCAGGGCGAG 508

Qy    557 GGGCCGATCCCGGTCCCTCCGTCGTCGCCCACTCAGCCTTCCGGCCTCTTCCACGGAGCCTT 616
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    509 GGGCCGATCCCGGTCCCTCCGTCGTCGCCCACTCAGCCTTCCGGCCTCTTCCACGGAGCCTT 568

Qy    617 AGCCATCACATCGACGTCAGACCTTCGGCCTCTCGGCCTCTTCCACGGAGCCTT 676
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    569 AGCCATCAACATCGACGTCAGACCTTCGGCCTCTCGGCCTCTTCCACGGAGCCTT 628

Qy    677 CTCCTGCAGGGCGAGAGCGTTCCTCCAGTTGTGGGTCGTGGTGCATCTCCCCTA 736
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    629 CTCCTGCAGGGCGAGAGCGTTCCTCCAGTTGTGGGTCGTGGTGGCATCTCCCCTA 688

Qy    737 AAGGAAGTGGTCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTTGCCTCCAGCTCTCCCAG 796
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    689 AAGGAAGTGGTCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTTGCCTCCAGCTCTCCCAG 748

Qy    797 ACAGAAAGACTCGTCCTTTTCGTCGTTGAACCCCAGGATCAGGAGGATTTGGAGCCCG 856
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    749 ACAGAAAGACTCGTCCTTTTCGTCGTTGAACCCCAGGATCAGGAGGATTTGGAGCCCG 808

Qy    857 TGAAGAAGAAATGAGAGAGATGGGGACCTTGACCTGAACGGGCAGTTGTTGGTCGCAC 916
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    809 TGAAGAAGAAATGAGAGAGATGGGGACCTTGACCTGAACGGGCAGTTGTTGGTCGCAC 868

Qy    917 AACCGGTAGAAATGCCCAAACCGTCCACGAGACCGTCAGAGCAGCAGCGGCTGGAAGCCAG 976
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    869 AACCGGTAGAAATGCCCAAACCGTCCACGAGACCGTCAGAGCAGCAGCGGCTGGAAGCCAG 928

Qy    977 ACAAGATGAGGAGACGCTGACAATGCATCATCTGCCAGGACCTGCTGCACGACTGCTGTGA 1036
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    929 ACAAGATGAGGAGACGCTGACATCATCATCTGCCAGGACCTGCTGCACGACTGCTGTGA 988

Qy   1037 GTTTGCAGCCCTGCATGCACGTTCTCGCGCGGTTCGGCGGCGGATCGTGAGCGGGCGCT 1096
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    989 GTTTGCAGCCCTGCATGCACGTTCTCGCGCGGTTCGGCGGCGGATCGTGATGAGCGGGCT 1048

Qy   1097 CGTCCCTGTCTCCTACCTGCGCGCTGCCCGTGGAGCGGGATCTGTAAAAACCACATCCTCA 1156
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1049 CGTCCCTGTCTCCTACCTGCGCGCTGCCCGTGGAGCGGGATCTGTAAAAACCACATCCTCA 1108

Qy   1157 ACAACCTGTGGAAGCATACCTCATCCAGACAAGAGTCGCAGTGAAGAAGATG 1216
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1109 ACAACCTGTGGAAGCATACCTCATCCAGACAAGAGTCGCAGTGAAGAAGATG 1168

Qy   1217 TGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGC 1276
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1169 TGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGC 1228

Qy   1277 GGTCTTTTCTGATGAAGAAGGGAGTTCAGAGGACCTGCTGGAGCGTGTCAGACGTTGACA 1336
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1229 GGTCTTTTCTGATGAAGAAGGGAGTTCAGAGGACCTGCTGGAGCGTGTCAGACGTTGACA 1288

Qy   1337 GTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTGCCCGCAGTGTCCTGAGTACAGAA 1396
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1289 GTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTGCCCGCAGTGTCCTGAGTACAGAA 1348

Qy   1397 GGCAGGCGGCCAGCCTCCCACTGCCCAGAGGGGGAGCCCCACAGG 1456
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1349 GGCAGGCGGCCAGCCTCCCACTGCCCAGAGGGGGAGCCCCACAGG 1408

Qy   1457 CCCTGGGGGATGCACCCTCCACGTCGTCGCAGCTTCAGCGTGACGACAGATTACGTGT 1516
```

```
Qy     17 GCCGCAATGTCTCTTGACACAGCGCGGCGGCCGGTTCGGCGGCGGGGCGG 76
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db      5 GCCGCAATGTCTCTTGACACAGCGCGGCGGCCGGTTCGGCGGCGGGGCGG 64

Qy     77 GGATGTGAATCCCGATGGAGCGGCCCGAGGAAGGCAAGCAGTGCGCCGCCCGCGCCCT 136
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65 GGATGTGAATCCCGATGGAGCGGCCCGAGGAAGGCAAGCAGTGCGCCGCCCGCGCCCT 124

Qy    137 GGGGACGGCTCCTGCGTCTGGGCGTCCTGGGCGTCCGCGGGCCGAGCCCGCACGTCCTCCTCGAGGAAGC 196
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 GGGGACGGCTCCTGCGTCTGGGCGTCCTGGGCGTCCGCGGGCCGAGCCCGCACGTCCTCCTCGAGGAAGC 184

Qy    197 GGGAGTGGACCATCGGGCGGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGG 256
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    185 GGGAGTGGACCATCGGGCGGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGG 244

Qy    257 TCTCTGAGATCACTGTAGAATTGTAGTGATGAATTGTAGTGGAACTGGAAG 316
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    245 TCTCTGAGATCACTGTAGAATTGTAGTGATGAATTGTAGTGGAACTGGAAG 304

Qy    317 ATACCAGCACCAGTGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGACAGATGCC 376
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    305 ATACCAGCACCAGTGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGACAGATGCC 364
```

```
Db   1469 GCCCTCTGCAAGGAAGCCACGCCGCCCTGTGCACCTGCTTCCAGCCATGCCGACCGGA 1528
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1577 GAGCGGAGCGGCGAGCAGGAGACCCGGTGTCGCCCTACGGCCAGTGTGCGGTCTGCTGCAGC 1636

Db   1529 GAGCGGAGCGGCGAGCAGGAGACCCGGTGTCGCCCTACGGCCAGTGTGCGGTCTGCTGCCTGCAGC 1588
          |||||||||||||||||||||||||||||||||||||||||
Qy   1637 CTTTCTGCCACCTGTACTGGGGCTGCACCCGGTCACCCGAGCCGGTGCTGCCTGGCCCCGT 1696

Db   1637 CTTTCTGCCACCTGTACTGGGGCTGCACCCGAGCCGGTGCTGCCTGGCCCCGT 1696
          |||||||||||||||||||||||||||||||||||
Qy   1589 CTTTCTGCCACCTGTACTGGGGCTGCACCCGACCCGGCTGCTACCGGCTGCCTGGCCCCGT 1648

Db   1697 TTTGTGAGCTCAACCTGGGTGACAAGTGTCTGGACGGCGGTCGACAACAGTACG 1756
          |||||||||||||||||||||||||||||||||||||||||||||
Qy   1649 TTTGTGAGCTCAACCTGGGTGACAAGTGTCTGGACGGCGTGACAACAGTACG 1708

Db   1757 AGTCAGACATCCTGAAGAATTACCTGCAACCAGAGTTTGACATGGAAAAACATGTTGA 1816
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy   1709 AGTCAGACATCCTGAAGAATTACCTGCAACCAGAGTTTGACATGGAAAAACATGTTGA 1768

Db   1817 CCGAGAGCCTCGTGGCCTCTCCAGCTCTCCCAG 1876
          |||||||||||||||||||||||||||||||
Qy   1769 CCGAGAGCCTCGTGGCCTCTCCAGCTCTCCCAG 1828

Db   1877 TTTGTGAGCTCAACCTGGGTGACAAGTGTCTGCGCAGCTTCCGTGAGCTGACCTATCAGT 1936
          |||||||||||||||||||||||||||||||||||||||||||
Qy   1829 GAGACACCGTTCGTTGTTACTGCGTGTGCCAGCTTCCGTGAGCTGACCTATCAGT 1888

Db   1937 ATCGGGCGCAGAACATTCCTGCTTCCGCACTCAGTGAAAGCTCACCACGCCATGAAATTCAATCATA 1996
          |||||||||||||||||||||||||||||||||||||||
Qy   1889 ATCGGGCAGAACATTCCTGCTTCCGCACTCAGTGAAAGCTCACCACGCCATGAAATTCAATCATA 1948

Db   1997 ACTGGGGCCGTAACTGCGCACTCAGTGAAAGCTCACCACGCCCTGACGCTTTCAGCAC 2056
          |||||||||||||||||||||||||||||||||||||
Qy   1949 ACTGGGGCCGTAACTGCGCACTCAGTGAAAGCTCACCACGCCCTGACGCTTTCAGCAC 2008

Db   2057 TCTGTGAACAGACAAGGTCAAAAACTAAGCATCCAGAGACATCGAGACAAGACAGCTTCAGCAC 2116
          |||||||||||||||||||||||||||||||||||||||
Qy   2009 TCTGTGAACAGACAAGGTCAAAAACTAAGCATCCAGAGACATCGAGACAAGACAGCTTCAGCAC 2068

Db   2117 TGGAGGTGAAGAGACCGTGTTTTTAAAATACAGAGACAAGAAGTGTTTTTCACA 2176
          |||||||||||||||||||||||||||||||||||
Qy   2069 TGGAGGTGAAGAGAGACCGTGTTTTTAAAATACAGAGACAAGAAGTGTTTTCACA 2128

Db   2177 GCCCCCTGAGGGAAGGACGCAGGGTCTCTGGGGTGACTCTTCTGTGG 2236
          ||||||||||||||||||||||||||||||||||||||
Qy   2129 GCCCCCTGAGGGAAGGACGCAGGGTCTCTGGGGTGACTCTTCTGTGG 2188

Db   2237 AGCTTTTTACCCTCGAGTGAGACCCCCCGAGCCCCGGAGCCGCCGCCCC 2296
          |||||||||||||||||||||||||||||||||||||
Qy   2189 AGCTTTTTACCCTCGAGTGAGACCCCCCGAGCCCCGGAGCCGCCGCCCC 2248

Db   2297 TGGTGAGCCGCTGGCAGGGCATCGGTGGCATCAGCAGCCCAAAGTCAGGC 2356
          ||||||||||||||||||||||||||||||||||||
Qy   2249 TGGTGAGCCGCTGGCAGGGCATCGGTGGCATCAGCAGCCCAAAGTCAGGC 2308

Db   2357 CATGCGGCCGTCCCGCGAGAGGGCAGTTTGTCTCTTTGTACATTTTCCGAAACTACA 2416
          ||||||||||||||||||||||||||||||||||||||
Qy   2309 CATGCGGCCGTCCCGCGAGAGGGCAGTTTGTCTCTTTGTACATTTTCCGAAACTACA 2368

Db   2417 GTTAAAGCAGAAGTCTGTTTTCAGGAAAGTTCAAGGGAGGAGGGCAAGTTTATCAAAA 2476
          ||||||||||||||||||||||||||||||||||||||
Qy   2369 GTTAAAGCAGAAGTCTGTTTTCAGGAAAGTTCAAGGGAGGAGGGCAAGTTTATCAAAA 2428

Db   2477 ACATTGTTTCAGGAGAAGGGAGCATAAGTTACAGCAGTTACAGCACAATATCCTG 2536
          ||||||||||||||||||||||||||||||||||||
Qy   2429 ACATTGTTTCAGGAGAAGGGAGCATAAGTTACAGCCTACAGCGTACACAATATCCTG 2488

Db   2537 CTGCTTGGGAAAACCACCAGCATTTATCATTTTATTTTTAATTATTTTAATAGGTTTGGTGCTTATCT 2596
          ||||||||||||||||||||||||||||||||||||
Qy   2489 CTGCTTGGGAAAACCACCAGCATTTATCATTTTATTTTTAATTATTTAATAGGTTTGGTGCTTATCT 2548

Db   2597 TCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATATAATTACAAA 2656
          |||||||||||||||||||||||||||||||||
Qy   2549 TCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATATAATTACAAA 2608
```

```
Db    425 TGGCATACCTCTATGAATCTTTAAGTGAATGAAAAGCAGGCATGACACAGAGAATCCTTT----- 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    497 CTAACAAGGAAAAATGTTGTTCCATGGGACCAAAGATACCTCAGGTGCAGGTGCAGGGGCAG 556

Db    481 --------GATACCTCAGGTGACGTCGAGGTGCAGGGGCAG 508
                 |||||||||||||||||||||||||||||
Qy    557 GGGCCGATCCCCGGGTCCCTCCGGTCGCCGGCACTCAGGTGCTTTGAGGAACCAC 616

Db    509 GGGCCGATCCCCGGGTCCCTCCGGTCGCCGGCACTCAGGTGCTGTTTGAGGAACCAC 568
          ||||||||||||||||||||||||||||||||||||||||
Qy    617 AGCCATCAACATGCAGCGTCAGACCTCTTCGGCCTCTTCCACGGAGCCTT 676

Db    617 AGCCATCAACATGCAGACGTCAGACCTCTTCGGCCTCTTCCACGGAGCCTT 676
          |||||||||||||||||||||||||||||||||||||||
Qy    569 AGCCATCAACGACGTCAGACCTCTTCCCACAGCCTCGGCCTCTTCCACGGAGCCTT 628

Db    677 CTCCTGCAGGCGGAGCGAGAGCGGTTCCTCCAGTTGTGGTCTGGGGTGGTGCATCCCCCTA 736
          |||||||||||||||||||||||||||||||||||||||
Qy    629 CTCCTGCAGGCGGAGCGAGAGCGTTCCTCCAGTTGTGGTCTGGGGTGGTGCATCCCCCTA 688

Db    737 AAGGAAGTGGTCCCCTCGTTGCGCCGGCCACTCAGCTCTTGCCTTCCAGCTCTCCCAG 796
          |||||||||||||||||||||||||||||||||||||||
Qy    689 AAGGAAGTGGTCCCCTCGTTGGCAAGTGATGATGAAGTCTCCAGCTTTGCCTTCCAGCTCTCCCAG 748

Db    797 ACAGAAAGACTGCGTCCTTTTCGTCGTTGGATGGAGATGGGGACCTTGACCTGGAACCAGAGGATTTGGAGCCCG 856
          |||||||||||||||||||||||||||||||||||||||
Qy    857 TGAAGAGAAATGAGAGGAGATGGGGGAGATCAAGGGCCAGTGATTGGTTGTCGCAC 916

Db    809 TGAAGAGAAATGAGAGGAGATGGGGGAGATCAAGGGCCAGTGATTGGTTGTCGCAC 868
          |||||||||||||||||||||||||||||||||||||||
Qy    749 ACAGAAAGACTGCGTCCTTTTCGTCGTTGGATGGGGGACCTTGACCTGGAACCCAGAGGATTTGGAGCCCG 808

Db    917 AACCCGTTAGAAATGCCCAAACCGTCCAAACCGTCAGCGAGACCGTCAGGAAGCCAG 976
          |||||||||||||||||||||||||||||||||||||||
Qy    869 AACCCGTTAGAAATGCCCAAACCGTCCAAACCGTCAGCGAGACAGCGGCGGGAAGCCAG 928

Db    977 ACAGATGGAGGAGACGCTGACACGTTCTGCGCGGCTTGCTACTCGCGGGCTGGATGGAGCGCT 1036
          |||||||||||||||||||||||||||||||||||||||
Qy    989 ACAAGATGGAGGAGACGCTGACACGTTCTGCGCGGCTTGCTACTCGCGGGCTGGATGGAGCGCT 1048

Db   1037 GTTTGCAGCCCTGCATGCACACGTTCTGCGCGGCTTGCTACTCGCGGGCTGGATGGAGCGCT 1096
          |||||||||||||||||||||||||||||||||||||||
Qy   1097 CGTCCCCTGTGTCTCGGCCGTGTCCCCGTGGAGCGGATCTGTAAAACCACATCCTCA 1156

Db   1097 CGTCCCCTGTGTCTCGCCGTGTCCCCGTGGAGCGGATCTGTAAAACCACATCCTCA 1156
          |||||||||||||||||||||||||||||||||||||||
Qy   1049 CGTCCCCTGTGTCCTGCCGTGTCCCCGTGGAGCGGATCTGTAAAAACCACTACCTCA 1108

Db   1157 ACAACCTCGTGGAAGCCATACCTCAGATCCAGACAAGAGTCGCAGTGAAGAAGATG 1216
          |||||||||||||||||||||||||||||||||||||||
Qy   1109 ACAACCTCGTGGAAGCATACCTCATCCAGAACAAGAGTCGCAGTGAAGAAGATG 1168

Db   1217 TGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGC 1276
          |||||||||||||||||||||||||||||||||||||||
Qy   1169 TGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGC 1228

Db   1277 GGTCTTTTTCTGATGAAGAGAAGGGAGTTCAGGAGGACCTGTCGTGTCCGGCAGTGTCCTGAGTTCAGAA 1336
          |||||||||||||||||||||||||||||||||||||||
Qy   1229 GGTCTTTTTCTGATGAAGAGAAGGGAGTTCAGGAGGACCTGTCGTGTCCGGCAGTGTCCTGAGTTCAGAA 1288

Db   1337 GTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTGCCCAGCACCCGAGGGCGAGCCCCCACAGG 1396
          |||||||||||||||||||||||||||||||||||||||
Qy   1349 GTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTGCCCAGCACCCGAGGGCGAGCCCCCACAGG 1408

Db   1397 GGCAGGCGGCGGCAGCGGCTCCCCACTGCCCACTGCCCCACTCGCCCACTCGCGCTCCCAGCACCCGAGGGCGAGCCCCCACAGG 1456
          |||||||||||||||||||||||||||||||||||||||
Qy   1457 CCCTGGGGATGCACCCTCAGCCTCGACGACAGCAGTCCAGGATTACGTGT 1516

Db   1457 CCCTGGGGATGCACCCTCAGCCTCGACGACAGCAGTCCAGGATTACGTGT 1516
          |||||||||||||||||||||||||||||||||||||||
Qy   1409 CCCTGGGGATGCACCCTCAGCCTCGACGACAGCAGTCCAGGATTACGTGT 1468

Db   1517 GCCCTCTGCAAGGAAGCCACGCCCGTGCACCTGCTTCCAGCCATGCCGACCGGA 1576
          |||||||||||||||||||||||||||||||||||||||
```

```
AX877194
LOCUS           AX877194                2639 bp    DNA        linear   PAT 17-DEC-2003
DEFINITION      Sequence 12099 from Patent EP1074617.
ACCESSION       AX877194
VERSION         AX877194.1  GI:40031930
KEYWORDS        .
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1
  AUTHORS       Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
                Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE         Primers for synthesising full-length cDNA and their use
  JOURNAL       Patent: EP 1074617-A 12099 07-FEB-2001;
                Research Association for Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1..2639
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             79..2037
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAE89614.1"
                     /db_xref="GI:40031931"
                     /translation="MERPEEGKQSPPPQPWGRLLRLGABEGEPHVLLRKREWTIGRRR
                     GCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDV
                     IYLVYRVRNEPEHVAYLYESLSEKQGMTQESPDTSGAGRGADPRVPPSSPATQVCF
                     EEPQPSTSTSDLFPTASASTEPSPAGRERSSSCGSGGGISPKGSGPSVASDEVSSF
                     ASALPDRKTASFSSLEFQODQEDLEHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV
                     RAAAGKPDKMEBTLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV
                     BRICKNHILNNLVEAYLIQHPDKSREEDVQSMDARNKITODMLQPKVRRSFSDEEGS
                     SEDLLELSDVDBSSDISQPYYVCRQCPEYRQAAQPPHCPAPEGEPGAPQALGDAPS
                     TSVSLTTAVQDYVCPLQGSHALCTCCPQPMDRRAEREQDPRVAPQQCAVCLQPFCHL
                     YWGCTRTGCYGCLAPFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTES
                     LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQRNIPASELPVAVTSRPDCY
                     WGRNCRTQVKAHAMKFNHICEQTRFKN"
ORIGIN

Query Match        95.8%;   Score 2556.4;   DB 6;   Length 2639;
Best Local Similarity    98.6%;   Pred. No. 0;
Matches 2613;   Conservative  0;   Mismatches  0;   Indels  36;   Gaps   1;

Qy       17 GCCGCAATGTCTCCTGACAGCGCGGCGGCGGCGGCGCGGTTCGGCGCGCGGGGGGCGG      76
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        5 GCCGCAATGTCTCCTGACAGCGCGGCGGCGGCGGCGCGGTTCGGCGCGCGGGGGGCGG      64

Qy       77 GGATGTGAATCCCATGGACAGCAGCAAGCAGTCCGCGGAGGAAGCAAGCAGTCCGCGCCGCAGCCCT     136
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       65 GGATGTGAATCCCATGGACAGCAGCAAGCAGTCCGCGGAGGAAGCAAGCAGTCCGCGCCGCAGCCCT     124

Qy      137 GGGGACGGCTCCTGCGTCTGGCGCGGGAGGGGCAGCCGCACGTCCTCCTCGAGGAAGC     196
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      125 GGGGACGGCTCCTGCGTCTGGCGCGGGAGGGGCAGCCGCACGTCCTCCTCGAGGAAGC     184

Qy      197 GGGAGTGGACCATCGGCGCGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGG     256
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      185 GGGAGTGGACCATCGGCGCGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGG     244

Qy      257 TCTCTGGAGATCACTGTAGAATTGTAGTGGAAGTTGTTAAGAACAAGCTGAAGGTTGTTAAGAGCAGCAC     316
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      245 TCTCTGGAGATCACTGTAGAATTGTAGTGGAAGTTGTTAAGAACAAGCTGAAGGTTGTTAAGAGCAGCAC     304

Qy      317 ATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAACAGCAGACATGCC     376
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      305 ATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAACAGCAGACATGCC     364

Qy      377 CTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAATGAACCGGAACACAACG     436
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      365 CTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAATGAACCGGAACACAACG     424

Qy      437 TGGCATACCCTCTATGAATCTTAAGTAAGAAAAGCAAGGCATGACACAAGAATCCTTTGAAG     496
```

```
Db       1621 GCGGTCCTGCCAGCCCTTTCTGCCACCTGTACTCGGGGCTCAACCTGGGCTGAGCTCAACCTGTCTGACACCGGACCGGCGTGCTG     1680
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       1681 GGCGTCCTGCCCGTTTGTGAGCTCAACCTGGGCTGAGCTCAACCTGTCTGACACGGCGTGCTG     1740
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GGCTGCCTGCCTGCCCGTTTGTGAGCTCAACCTGGGCTGACAAGTGTCTGACGGCGTGCTG     1740

Qy       1741 AACAACAACAGCTACGAGTCAGACATCCTGAAGAATTACCTGGCAACCAGAGGTTTGACA     1800
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 AACAACAACAGCTACGAGTCAGACATCCTGAAGAATTACCTGGCAACCAGAGGTTTGACA     1800

Qy       1801 TGGAAAAACATGTTGACCGAGCCTCGTGGCCTCTCCAGCGGGGAGTGTTTCTGCTGTCT     1860
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TGGAAAAACATGTTGACCGAGCCTCGTGGCCTCTCCAGCGGGGAGTGTTTTCTGCTGTCT     1860

Qy       1861 GATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGTGTGCGCCTGCGCAGCTTCCGT     1920
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGTGTGCGCCTGCGCAGCTTCCGT     1920

Qy       1921 GAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGGCCGTAACA     1980
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGGCCGTAACA     1980

Qy       1981 TCCCGTCCGTACTGCGACTGCTACTGGGGCCGTAACTGCCGTCAGGTGAAAGCTCACCACGCC     2040
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TCCCGTCCGTACTGCGACTGCTACTGGGGCCGTAACTGCCGTCAGGTGAAAGCTCACCACGCC     2040

Qy       2041 ATGAAATTCAATCAATATCGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCT     2100
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 ATGAAATTCAATCAATATCGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCT     2100

Qy       2101 GAGCAGCTTTCAGCACTGGAGAGTGAAGAGAGCGTGTGTTTTAAAATACAGACAAGCACG     2160
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GAGCAGCTTTCAGCACTGGAGAGTGAAGAGAGCGTGTGTTTTAAAATACAGACAAGCACG     2160

Qy       2161 TCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGGACGCAGGTCTCCGACAGTGCTCTGG     2220
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 TCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGGACGCAGGTCTCCGACAGTGCTCTCTG     2220

Qy       2221 GGTGACTCCTTCTGTGTGGAGCTTTTACCCTCTGAGTGAGACCCTCCCAGAGCCCCGGGGG     2280
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 GGTGACTCTCTGTGTGGAGCTTTTACCCTCTGAGTGAGACCCTCCCAGAGCCCCGGGGG     2280

Qy       2281 CCGCAGCCCGCCCTCCTGGGTGGCGCTGGGCAGGGCTGCGTGTGGCATCAGCAGCAGAGA     2340
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 CCGCAGCCCCGCCCCTCCTGGGTGGCGCTGGGCAGGGCTGCGTGTGGCATCAGCAGCAGAGA     2340

Qy       2341 CGAAGCCTTTCTGTAACATGGGACCGTCCCGCCCGGAGAGGGGCAGTTTGCTCTTTTGTAC     2400
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CGAAGCCTTTCTGTAACATGGGACCGTCCCGCCCGGAGAGGGGCAGTTTGCTCTTTTGTAC     2400

Qy       2401 ATTTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTCAGGAAAAGTTTCAAGGGAGAAG     2460
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 ATTTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTCAGGAAAAGTTTCAAGGGAGAAG     2460

Qy       2461 GGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGAGCATAAGTTTACAGCCTACACGGA     2520
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 GGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGAGCATAAGTTTACAGCCTACACGGA     2520

Qy       2521 CGTACCAAATATCCTGCTCGTGCTGGGAAAACCACAGCATTTTATCTATTTTTTATTTTAATA     2580
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 CGTACCAAATATCCTGCTCGTGCTGGGAAAACCACAGCATTTTATCTATTTTTTATTTTAATA     2580

Qy       2581 GGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATATA     2640
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 GGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATATA     2640

Qy       2641 ATTTATAATTTACAAATTGACAAAAAAAAAAAAAAAAAAAAA     2679
                  ||||||||||||||||||||||||||||||||||
Db       2641 ATTTATAATTTACAAATTGACAAAAAAAAAAAAAAAAAAAAA     2679
```

RESULT 2

FEATURES   Location/Qualifiers
     source   1..2679
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS   91..2085
                     /note="contains FHA domain and ring finger"
                     /codon_start=1
                     /product="cell cycle checkpoint protein CHFR"
                     /protein_id="AAF91084.1"
                     /db_xref="GI:9651170"
                     /translation="MERPEGKQSPPPQPWGRLLRLGABEGEPHVLLRKREWTIGRRR
                     GCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVAKQTCPLQTGDV
                     IYLVYRKNEPEHNVAYLIESLSEKQGMTQESFRANKENVFHGTKDTSGAGARGADPR
                     VPPSPATQVCFEEPQPSTSTSDLFPTASASSTRPSPAGRERRSSCSGGGGISPKGS
                     GPEVASDEVSSPASALPDRKTASFSSLEPQODEDLEPVKKKMRGDDLDLNGQLLVAQ
                     PRRNAQTVHEDVRAAAGKPKDKMEFTLTCIICQDLLHDCVSLQPCMHTFCAACYSGWME
                     RSSLCPTCRCPVERICKNHILNNIVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQP
                     KVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPHCPAPEGE
                     PGAPQALGDAPSTSVSLITAVQDYVCPLQGSHALCTCCFQPMDRRAEREQDPRVAPQ
                     QCAVCLQPFCHLYRGCTRTGYGCLAPFCELNLGDKCHLDGVLNNSYRSDILNYLAT
                     RGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFPRELTYQRQNIPASE
                     LPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQTRFKN"

ORIGIN

Query Match   100.0%;   Score 2679;   DB 9;   Length 2679;
Best Local Similarity   100.0%;   Pred. No. 0;
Matches 2679;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 AAGAATTCGGCACGAGGCCGCCAATGTCTCTTGACACGGCGGCGGCGGGCGGCCGTTCCGG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AAGAATTCGGCACGAGGCCGCCAATGTCTCTTGACACGGCGGCGGCGGGCGGCCGTTCCGG 60

Qy   61 GTTCGGCGCGGCGGGGCGGGGATGGAATCGGCGGGGGCGCCGAGGAAGCGAGCCAGCAGT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GTTCGGCGCGGCGGGGCGGGGATGGAATCGGCGGGGGCGCCGAGGAAGCGAGCCAGCAGT 120

Qy   121 CGGCCGCCGCAGCCGTTGGGGACGGGCCTCCTGCGTCTCGGGCGGCGGAGGAGGCCGCAC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CGGCCGCCGCAGCCGTTGGGGACGGGCCTCCTGCGTCTCGGGCGGCGGAGGAGGCCGCAC 180

Qy   181 GTCCTCCTGAGGAAGCGGGAGTGGACCATCGGGCCGGAGACGAGGTTGCGACCTTTCCTTC 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTCCTCCTGAGGAAGCGGGAGTGGACCATCGGGCCGGAGACGAGGTTGCGACCTTTCCTTC 240

Qy   241 CCCAGCAATAAACTGTCTCTGGAGATCACTGTAGAATTGTAGTGCGATGAAAAATCAGGT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCCAGCAATAAACTGTCTCTGGAGATCACTGTAGAATTGTAGTGCGATGAAAAATCAGGT 300

Qy   301 CAGGTGACACTGGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CAGGTGACACTGGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTT 360

Qy   361 AAGAAGCAGACATGCCCTTTACAGACTGGGGATGTCATCTCACTACTTGGTGTACAGGAAGAAT 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGAAGCAGACATGCCCTTTACAGACTGGGGATGTCATCTCACTACTTGGTGTACAGGAAGAAT 420

Qy   421 GAACCGGAACAACAACGTGGCATAACCTTAAGTGAAAAGCAAGGCATGACA 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAACCGGAACAACAACGTGGCATAACCTTAAGTGAAAAGCAAGGCATGACA 480

Qy   481 CAAGAATCCTTTGAAGCTAACAAGGAAAAATGTGTTCCATGGACCAAAGATACCTCAGGT 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CAAGAATCCTTTGAAGCTAACAAGGAAAAATGTGTTCCATGGACCAAAGATACCTCAGGT 540

Qy   541 GCAGGTGCAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGATCCCCGGGTCCTCCTCGTCGTCCGCCGCCACTCAGGTG 600
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GCAGGTGCAGGGGCGGAGGGGCGGAGGGGCGGAGGGGCGGATCCCCGGGTCCTCCTCGTCGTCCGCCGCCACTCAGGTG 600

Qy   601 TGCTTGAGGAACCACAGCCATCAACATCGAGCTCAGACCTCAGAGCTCGGCC 660
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGCTTGAGGAACCACAGCCATCAACATCGAGCTCAGACCTCAGAGCTCGGCC 660

Qy   661 TCTTCCACGGAGCCTTCCTCCTGCAGGGGGGAGGAGGGTTCCTCCAGTTGTGGGTCTGGGGGT 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TCTTCCACGGAGCCTTCCTCCTGCAGGGGGGAGGAGGGTTCCTCCAGTTGTGGGTCTGGGGGT 720

Qy   721 GGTGGCATCTCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAAGTCTTCCAGCTTT 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GGTGGCATCTCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAAGTCTTCCAGCTTT 780

Qy   781 GCCTCAGCTCTCCCAGACAGAAAGACTGGTCCTTTGTCGTTGGAACCCCAGGATCAG 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCCTCAGCTCTCCCAGACAGAAAGACTGGTCCTTTGTCGTTGGAACCCCAGGATCAG 840

Qy   841 GAGGATTGGAGCCGTGAAGAGAAAATGAGAGGATGGGACCTTGACCTGAACGGG 900
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 GAGGATTGGAGCCGTGAAGAGAAAATGAGAGGATGGGACCTTGACCTGAACGGG 900

Qy   901 CAGTTGTTGGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAGACGTCAGAGCA 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 CAGTTGTTGGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAGACGTCAGAGCA 960

Qy   961 GCGGCTGGGAAGCCAGACAAGATCGAGGAGACGCTGACATGCATCATCTGCCAGGACCTG 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GCGGCTGGGAAGCCAGACAAGATCGAGGAGACGCTGACATGCATCATCTGCCAGGACCTG 1020

Qy   1021 CTGCACGACTGCGTGAGTTGCAGCCCTGCATGCACGTTCTGCGCGGCTTGCTACTCG 1080
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CTGCACGACTGCGTGAGTTGCAGCCCTGCATGCACGTTCTGCGCGGCTTGCTACTCG 1080

Qy   1081 GGCTGGATGGAGCGCTCGTCCTGTCCTGCTGGAAGCCGGCGGCGGGCGGTTCCGG 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GGCTGGATGGAGCGCTCGTCCTGTCCTGCTGGAAGCCGGCGGCGGGCGGTTCCGG 1140

Qy   1141 AAAAACCCATCTTCAACAACCTCGTGGAAGCATACCTCATCCAGCATCCAGACAAAGAGT 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AAAAACCCATCTTCAACAACCTCGTGGAAGCATACCTCATCCAGCATCCAGACAAAGAGT 1200

Qy   1201 CGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTG 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTG 1260

Qy   1261 CAGCCCAAAGTCAGGCGGTCTTTTTCTGATGAAGAAGGGAGTTCAGAGGACCTTGCTGGAG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CAGCCCAAAGTCAGGCGGTCTTTTTCTGATGAAGAAGGGAGTTCAGAGGACCTTGCTGGAG 1320

Qy   1321 CTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAACATTAGCCAGCCATAGTCGTGTGCCGGCAG 1380
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAACATTAGCCAGCCATAGTCGTGTGCCGGCAG 1380

Qy   1381 TGTCCTGAGTACAGAAGGCAGGCGGCCAGCCTCCCCACTGCCCAGAGGGCAG 1440
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TGTCCTGAGTACAGAAGGCAGGCGGCCAGCCTCCCCACTGCCCAGAGGGCAG 1440

Qy   1441 CCAGGAGCCCCACAGGGCCTGGGGGATGCACCCTCCAGGTCCGTCGACGACAGCA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CCAGGAGCCCCACAGGGCCTGGGGGATGCACCCTCCAGGTCCGTCGACGACAGCA 1500

Qy   1501 GTCCAGGATTACGTGTGCCCTCTGCAAGGAAGCCACGCCCTGTGCACCTGCTGCTTCCAG 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GTCCAGGATTACGTGTGCCCTCTGCAAGGAAGCCACGCCCTGTGCACCTGCTGCTTCCAG 1560

Qy   1561 CCCATGCCCGACCGGAGCGGAGGCGGAGGCGGAGGCGGAGGCAGCCGGTGTCGCCCCTCAGCAGTTGT 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CCCATGCCCGACCGGAGCGGAGGCGGAGGCGGAGGCGGAGGCAGCCGGTGTCGCCCCTCAGCAGTTGT 1620

Qy   1621 GCGGTCTGCCTGCCACCTGTTCTGCCACCTGTACTGGGGCTGCACCCGGACCGGCTGCTAC 1680

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: May 14, 2004, 11:30:48 ; Search time 16426 Seconds
(without alignments)
7069.036 Million cell updates/sec

Title: US-10-048-046-1
Perfect score: 2679
Sequence: 1 aagaattcggcacgaggccg........acaaaaaaaaaaaaaaaa 2679

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	2679	100.0	2679	9	AF170724	AF170724 Homo sapi
2	2566.4	95.8	2639	6	AX877194	AX877194 Sequence
3	2566.4	95.8	2639	6	BD156534	BD156534 Primer fo
4	2566.4	95.8	2639	6	AK027687	AK027687 Homo sapi
5	2497.2	93.2	3189	6	BC012072	BC012072 Homo sapi
6	2262.4	84.4	3138	9	AX877222	AX877222 Sequence
7	2262.4	84.4	3138	6	BD156548	BD156548 Primer fo
8	2262.4	84.4	3138	9	AK001658	AK001658 Homo sapi
9	2244.6	83.8	3181	6	AX405642	AX405642 Sequence
10	1919.4	71.6	2448	6	AX834949	AX834949 Sequence
11	1919.4	71.6	2448	9	AK097671	AK097671 Homo sapi
12	1500.4	56.0	2297	9	HSM802304	AL137561 Homo sapi
13	1491	55.7	3194	10	BC049792	BC049792 Mus muscu
14	1279.8	47.8	6235	6	BD183291	BD183291 Novel gen
15	961.8	35.9	2257	6	AK090948	AK090948 Homo sapi
16	876.2	32.7	2186	6	AX713476	AX713476 Sequence
17	876.2	32.7	2186	9	AK054917	AK054917 Homo sapi
18	727	27.1	816	6	AX868758	AX868758 Sequence
19	727	27.1	816	6	BD148820	BD148820 Primer fo
20	608.4	22.7	181012	9	AC127070	AC127070 Homo sapi
21	608.4	22.7	181438	9	AC023047	AC023047 Homo sapi
22	583	21.8	824	6	AX868776	AX868776 Sequence
23	583	21.8	824	6	BD148838	BD148838 Primer fo
24	488.8	18.2	518	6	AX873789	AX873789 Sequence
25	488.8	18.2	518	6	BD153851	BD153851 Primer fo
26	440.8	16.5	151210	2	AC139591	AC139591 Rattus no
27	440.8	16.5	273812	2	AC106441	AC106441 Rattus no
28	430.4	16.1	236751	2	AC128407	AC128407 Rattus no
29	416	15.5	61779	2	AC126301	AC126301 Rattus no
30	218	8.1	181438	2	AC023047	AC023047 Homo sapi
31	142.4	5.3	202792	2	AC125898	AC125898 Rattus no
32	139.2	5.2	158897	9	AC118260	AC118260 Mus muscu
33	93.4	3.5	173911	9	AC092490	AC092490 Homo sapi
34	85.2	3.2	338116	2	AC087159	AC087159 Mus muscu
35	82.8	3.1	321708	2	AC087142	AC087142 Mus muscu
36	82.2	3.1	617	6	AX385952	AX385952 Sequence
37	69.4	2.6	601	11	BV002855	BV002855 S20P6639
38	64.4	2.4	165089	9	AC021646	AC021646 Homo sapi
39	55.6	2.1	125020	9	AF429315	AF429315 Homo sapi
40	55.4	2.1	14923	3	AE001394	AE001394 Plasmodiu
41	54.8	2.0	125020	9	AF429315	AF429315 Homo sapi
42	53.4	2.0	61052	2	AC123513	AC123513 Dictyoste
43	53.4	2.0	136240	3	AC117070	AC117070 Dictyoste
44	52.6	2.0	110000	3	AC116957_2	Continuation (3 of
45	52.4	2.0	2000	6	AX655393	AX655393 Sequence

ALIGNMENTS

RESULT 1
AF170724
LOCUS   AF170724   2679 bp   mRNA   linear   PRI 03-AUG-2000
DEFINITION Homo sapiens cell cycle checkpoint protein CHFR mRNA, complete cds.
ACCESSION   AF170724
VERSION   AF170724.1 GI:9651169
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2679)
AUTHORS   Scolnick,D.M. and Halazonetis,T.D.
TITLE   Chfr defines a mitotic stress checkpoint that delays entry into
            metaphase

Pred. No. is the number of results predicted by chance to have a

Query Match   8.7%;   Score 82.5;   DB 5;   Length 1023;
Best Local Similarity   29.7%;   Pred. No. 4.1;
Matches   22;   Conservative   8;   Mismatches   27;   Indels   17;   Gaps   1;

```
Qy     31 APQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCB------------------LNLGDKCLDG 73
             ||| || :|||:|  |:|  || |  |||         |||  :  |||  ||||  :
Db    950 APNSCMGCKQPFTLLHVMKARVHCHCCGKIFCESCTSHKCPIKKYRINTPVRVCDKCFND 1009

Qy     74 VLNNNSYESDILKN 87
             |:|:| ||| ||:
Db   1010 LQSNPSSNSFIIND 1023
```

RESULT 15
Q8I0G8
ID   Q8I0G8;   PRELIMINARY;   PRT;   523 AA.
AC   Q8I0G8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG32018-PG (Zyx102.6 isoform).
GN   CG32018 OR ZYX102EF.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Faffan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Stronach B.E., Siegrist S., Renfranz P.J., Macalma T., Beckerle M.C.;
RT   "The Zyx102 gene of Drosophila melanogaster.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003846; AAN06571.1; -.
DR   EMBL; AF219947; AAN85868.1; -.
DR   InterPro; IPR001781; LIM.
DR   Pfam; PF00412; LIM; 3.
DR   ProDom; PD000094; LIM; 3.
DR   SMART; SM00132; LIM; 3.
DR   PROSITE; PS00478; LIM_DOMAIN_1; 2.
DR   PROSITE; PS50023; LIM_DOMAIN_2; 1.
SQ   SEQUENCE 523 AA; 59008 MW; 79B2A099F8D7890 CRC64;

Query Match   8.6%;   Score 81.5;   DB 5;   Length 523;
Best Local Similarity   24.9%;   Pred. No. 2.4;
Matches   44;   Conservative   17;   Mismatches   63;   Indels   53;   Gaps   10;

```
Qy      7 HALCTCC------FQPMP---DRRAREEQDPRVAPQQCAVCLQPFCHLYWGCT----RT 52
            |||:||        |  |  ::       |:||:| :|||::||:|||:||
Db    348 HIFCFTCTDCQINLQGKPFVALDGKPYCEYDYLQTLEKCSVCMEPILERILRATGKPYHP 407

Qy     53 GCYGCLAPFCEINLGDKCLDGVL-----NNSYESDILKNYLATRGLTWKNML------- 100
            || ||| |||:|||| |:|  |     :|||:|| | ||:  | :|:
Db    408 QCFTCVV--C----GKSLDGLLFTVDATNQNYCITDFHKKF-APRCCVCKQPIMPDPGQ 459

Qy    101 --TESLVALQR-----------GVFLLSDYRVTG----DTVLCYCCGLRSFRELT 138
                ::  | |            |:  |  |::| :    :  :|:||| |::|:|
Db    460 EETIRVALDRSFHLECYKCEDCGLLLSSEAEGRGCYPLDDHVLCKSCNAKRVQALT 516
```

Search completed: May 7, 2004, 14:49:06
Job time : 26.0148 secs

```
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00008; EGF; 7.
DR   Pfam; PF02494; HYR; 1.
DR   Pfam; PF00057; ldl_recept_a; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 3.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00032; CCP; 3.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00181; EGF; 13.
DR   SMART; SM00192; LDLa; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 8.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50068; LDLRA_2; 1.
KW   EGF-like domain.
SQ   SEQUENCE 2972 AA; 329457 MW; 87D7BA80562F4C74 CRC64;

Query Match        8.9%;  Score 83.5;  DB 5;  Length 2972;
Best Local Similarity  27.1%;  Pred. No. 11;
Matches  55;  Conservative  12;  Mismatches  51;  Indels  85;  Gaps  15;

Qy    10 CTCCFQP------MPDRRAEREQDPRVAPQQCAVCLQ--------------------P   41
            | |:||                :|
Db  1407 CLCNPQPPCPSTELLPQR-DQIGCTLNSCQNGVCLQMNGMHDCACNEYFIVDRENGTP 1464

Qy    42 FCH---LYWGCTRTGCYGC--------LAPFCELN-LGDKC-LDGVLNNNSYE---SD   83
            ||            :||         |||                       ||
Db  1465 FCKPNHCLFTNQTRNSGYDCSTGRENALFLCP--NLNEFGDFCQYEGQLYNNSYTYIlSA 1522

Qy    84 ILKNYLATRGLTWKNMLTESIVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQ 143
            ::| :|||      |||       :||
Db  1523 GIENAIATN-------LCESLDTL------YAVPN----TFCV-------------Q 1549

Qy   144 NIPASELPVAVTSRPDCYWGRNC 166
            ||      ::
Db  1550 N-PTSTTP-SIHRCDFCYGGENC 1570

RESULT 13
O16004
ID   O16004            PRELIMINARY;      PRT;   2531 AA.
AC   O16004;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Notch homolog.
OS   Lytechinus variegatus (Sea urchin).
OC   Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC   Echinoidea; Euechinoidea; Echinacea; Temnopleuroida; Toxopneustidae;
OC   Lytechinus.
OX   NCBI_TaxID=7654;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97454256; PubMed=9310331;
RA   Sherwood D.R., McClay D.R.;
RT   "Identification and localization of a sea urchin Notch homologue:
RT   insights into vegetal plate regionalization and Notch receptor
RT   regulation.";
RL   Development 124:3363-3374(1997).
DR   EMBL; AF000634; AAB82088.1; -.
DR   HSSP; P01132; 1EGF.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0030154; P:cell differentiation; IEA.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
```

```
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 34.
DR   Pfam; PF00066; notch; 3.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 5.
DR   SMART; SM00179; EGF_CA; 23.
DR   SMART; SM00004; NL; 3.
DR   PROSITE; PS50088; ANK_REPEAT; 5.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 21.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 25.
DR   PROSITE; PS01187; EGF_CA; 20.
DR   PIRSF; PIRSF002279; Notch; 1.
KW   ANK repeat; EGF-like domain; Repeat.
SQ   SEQUENCE 2531 AA; 273982 MW; 5BF42BEC627CA303 CRC64;

Query Match        8.8%;  Score 83;  DB 5;  Length 2531;
Best Local Similarity  28.6%;  Pred. No. 11;
Matches  30;  Conservative  9;  Mismatches  40;  Indels  26;  Gaps  6;

Qy     5 GSHALCTCCFQP--MPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTR-TGCYGCL---  58
Db   840 GYSCRCHSGFQGNFCDDDRNE-----------CLFSPCRNGGSCTNLEGSFECSCLP   885

Qy    59 ---APFCELNLGDKCLDGVLNNSYESDILKNYLAT-RGLTWKN  98
Db   886 GYDGPICEINI-DECASGPCTNGGICTDLIDDYFCSCQRGFTGKN 929

RESULT 14
Q9XYD4
ID   Q9XYD4            PRELIMINARY;      PRT;   1023 AA.
AC   Q9XYD4;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Sterol glucosyltransferase (EC 2.4.1.173).
GN   UGT52.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=99240683; PubMed=10224056;
RA   Warnecke D., Erdmann R., Fahl A., Hube B., Muller F., Zank T.,
RA   Zahringer U., Heinz E.;
RT   "Cloning and functional expression of UGT genes encoding sterol
RT   glucosyltransferases from Saccharomyces cerevisiae, Candida albicans,
RT   Pichia pastoris, and Dictyostelium discoideum.";
RL   J. Biol. Chem. 274:13048-13059(1999).
DR   EMBL; AF098916; AAD28546.1; -.
DR   GO; GO:0016306; F:sterol 3-beta-glucosyltransferase activity; IEA.
DR   GO; GO:0016758; F:transferase activity, transferring hexosyl . . .; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   GO; GO:0030259; P:lipid glycosylation; IEA.
DR   InterPro; IPR004276; Glyco_trans_28.
DR   InterPro; IPR004182; GRAM_dom.
DR   InterPro; IPR000306; Znf_FYVE.
DR   Pfam; PF01363; FYVE; 1.
DR   Pfam; PF03033; Glyco_transf_28; 1.
DR   Pfam; PF02893; GRAM; 2.
DR   SMART; SM00064; FYVE; 1.
DR   SMART; SM00568; GRAM; 1.
DR   PROSITE; PS50178; ZF_FYVE; 1.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE 1023 AA; 114366 MW; 0E3C8C556C5D3709 CRC64;
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=Friedlin;
RA   Aert R., Robben J., Volckaert G., Ivens A.C., Quail M.,
RA   Rajandream M.A., Barrell B.G.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Friedlin;
RX   MEDLINE=98146435; PubMed=9477341;
RA   Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA   Smith D.F.;
RT   "A physical map of the Leishmania major Friedlin genome.";
RL   Genome Res. 8:135-145(1998).
DR   EMBL; AL359774; CAB95236.2; -.
DR   GO; GO:0005489; F:electron transporter activity; IEA.
DR   GO; GO:0004806; F:triacylglycerol lipase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0006118; P:electron transport; IEA.
DR   GO; GO:0006629; P:lipid metabolism; IEA.
DR   InterPro; IPR006058; 2Fe2S_fd_BS.
DR   InterPro; IPR002921; Lipase_3.
DR   InterPro; IPR000306; Znf_FYVE.
DR   Pfam; PF01764; Lipase_3; 1.
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
DR   PROSITE; PS50178; ZF_FYVE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   806 AA;  88248 MW;  17E4A8B623E5DF7D CRC64;

Query Match             9.4%;  Score 88.5;  DB 5;  Length 806;
Best Local Similarity   21.0%;  Pred. No. 0.65;  Indels 71;  Gaps 8;
Matches   34;  Conservative   19;  Mismatches   38;

Qy    2 PLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPF--------CHLYWGCTRTG 53
                                                                       
Db   78 PLEG-HRVCTTC--------QRAKEGCIC-TTCSVCRSRFGFHANRHHCRRCWRAVCSG 126

Qy   54 CYGCLAPFCELNLGDKCLDGVLNNSYESDILKNYLATRG-------LTW--KNMLT 101
                                                                      
Db  127 CWG------------------------------YRHYVHTLGRQMMVCDRCSTLWALANLAT 158

Qy  102 BSLVALQRGVFLL----------SDYRVTGDTVLCYCCGL 131
                                                       
Db  159 RSDAGLLWGLYILRTVGPMPRMCIGSSCRILTRRLVCYGCGL 200

RESULT 11
Q9UPQ4
ID   Q9UPQ4        PRELIMINARY;      PRT;   504 AA.
AC   Q9UPQ4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein KIAA1098 (Fragment).
GN   KIAA1098.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   TISSUE=Brain;
RC   
RX   MEDLINE=9939752; PubMed=10470851;
RA   Kikuno R., Nagase T., Ishikawa K., Hirosawa M., Miyajima N.,
RA   Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XIV.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:197-205(1999).
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AB029021; BAA83050.1; -.
DR   Genew; HGNC:16285; TRIM35.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
```

```
DR   InterPro; IPR001870; B302.
DR   InterPro; IPR006574; PRY.
DR   InterPro; IPR003877; SPRY_receptor.
DR   InterPro; IPR000315; Znf_Bbox.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00622; SPRY; 1.
DR   Pfam; PF00643; zf-B_box; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00336; BBOX; 1.
DR   SMART; SM00589; PRY; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50119; ZF_BBOX; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   504 AA;  57735 MW;  E0B377F32A36D1B9 CRC64;

Query Match             9.0%;  Score 85;  DB 4;  Length 504;
Best Local Similarity   24.8%;  Pred. No. 0.91;
Matches   34;  Conservative   15;  Mismatches   58;  Indels 30;  Gaps 7;

Qy   16 PMPDRRA-BREQDPRVAPQQ--------CAVCLQPFCH-LYWGCTRTGCYGCLAPFCELNL 66
                                         |||:||| ::: :
Db    5 PEPAARAMERSPDVSPGFSRSPFKEELLCAVCYDPFRDAVTLRCHNFCRGCVSRCWEVQV 64

Qy   67 GDKCLDGVLNNSYESDILKNYL------------ATRGLTWKOMLTESLVALQRG---VF 112
                                               ::| :
Db   65 SPTC--PVCKDRASPADLRTNHTLNNLVEKLLRBEAEGARWTSYRFSRVCRLHRGQLSLF 122

Qy  113 LLSDYRVTGDTVLCYCC 129
                               
Db  123 CLED-----KELLCCSC 134

RESULT 12
P90891
ID   P90891        PRELIMINARY;      PRT;  2972 AA.
AC   P90891;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   F55H12.3 protein.
GN   F55H12.3
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dobson R.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z81091; CAB03143.2; -.
DR   PIR; T22759; T22759.
DR   HSSP; P00740; 1EDM.
DR   WormPep; F55H12.3; CE25008.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003410; HyaTin.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR002172; LDL_receptor_A.
```

```
Query Match                    95.5%;  Score 901; DB 11; Length 663;
Best Local Similarity          93.4%;  Pred. No. 3.2e-92;
Matches  155;  Conservative   9;  Mismatches   2;  Indels   0;  Gaps   0;

Qy    1 CPLQGSHALCTCCFQPMPDRRAAREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  475 CPLQGSHALCTCCFQPMPDRRAAREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCFCCLAP 534

Qy   61 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 120
          |||||||||||||||| ||||||||||||||||||||: :||||| ||||||||:||||:
Db  535 FCELNLGDKCLDGVLNNNYESDILKNYLATRGLTWKSVLTESLLALQRGVFMLSDYRIT 594

Qy  121 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
          |:||||||||||||||||||||||||||||||||:|||||||||||
Db  595 GNTVLCYCCGLRSFRELTYQYRQNIPASELPVTVTSRPDCYWGRNC 640
```

RESULT 8
Q9SX88
ID   Q9SX88   PRELIMINARY;   PRT;   473 AA.
AC   Q9SX88;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   F16N3.15.
GN   F16N3.15.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. columbia;
RA   Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA   Lee J., Liu A., Li J., Kremenetskaia I., Gonzalez A., Altafi H.,
RA   Araujo R., Brooks S., Buehler E., Chao Q., Conn L., Conway A.B.,
RA   Dunn P., Hansen N., Huizar L., Khan S., Kim C., Palm C., Rowley D.,
RA   Shinn P., Walker M., Davis R.W., Ecker J.R., Federspiel N.A.,
RA   Theologis A.;
RT   "Arabidopsis thaliana chromosome 1 BAC F16N3 sequence.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AC007519; AAD46042.1; -.
DR   PIR; C96516; C96516.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE   473 AA; 53874 MW; D17217493A81C991 CRC64;

```
Query Match                    14.6%;  Score 137.5; DB 10; Length 473;
Best Local Similarity          25.0%;  Pred. No. 1.1e-06;
Matches   45;  Conservative  26;  Mismatches  88;  Indels  21;  Gaps   6;

Qy    1 CPLQGSHALCTCCFQPMPDRRAAREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
          | :||||| |||:|  ::    |         :   |  :
Db  275 CEHHGAHLQCHLCQQMPFRANLQ-----VPLHCKGCDRPFCGAYWS-SENVTQGVSGP 327

Qy   61 FCELN---LGDKCLDGV---LNNSYESDILKNYLATRGLTWKNMLTESL----VAL 107
          :|||:    ::::|  |     :  ::  |
Db  328 VCVRETFRPISERTITRIPFITHEMNRHEQDITQRCIAFMEKTVPDVVAEWLRLFNNREI 387

Qy  108 QRGVFLLSDYR-VTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
          :| :  :  :   :  : : |  :  :   :  |:::   ::: |
Db  388 DRSRMPLNHAETITASTHVCNDCYDKLVGFLLYWFRITLPRNHLPADVAAREDCWYGYAC 447
```

RESULT 9
Q7T3F2
ID   Q7T3F2   PRELIMINARY;   PRT;   292 AA.
AC   Q7T3F2;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=22338257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Strausberg R.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC053142; AAH53142.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   292 AA; 31678 MW; 5BBA8E434A66287C CRC64;

```
Query Match                     9.7%;  Score 91.5; DB 13; Length 292;
Best Local Similarity          23.2%;  Pred. No. 0.089;
Matches   35;  Conservative  20;  Mismatches  67;  Indels  29;  Gaps   7;

Qy    1 CPLQGSHALCTCCFQPMDRRAAREQDDPRVAP---------QQCAVCLQPFCHLY---- 46
          ||:|  : :||:         ::                   : | :: |
Db  125 CPVCQT-VFCCCCRSPWSDGHSCSLQQPLMPLPDGALSSDTEAAIKQCPVCGVIERN 183

Qy   47 WGCTRTGCYGCLAPFCELNLGDKCLDGVLNNNSYESDILKNYL--ATRGLTWKN------ 98
          :||   |   :    |  |  :   | :  : |:     :      :
Db  184 QGCAQMLCKSCKHTFCWVCLQN--LDGDIFLRHYDKGPCRNKLGHSRASVMNRTQVVGI 241

Qy   99 MLTESLVALQRGVFLLSDYRVTGDTVLCYCC 129
          :|  :  |  |  ::  :    :    | : |
Db  242 LVGASIVLVTSPLLL----LASPCIICCVC 268
```

RESULT 10
Q9N9P8
ID   Q9N9P8   PRELIMINARY;   PRT;   806 AA.
AC   Q9N9P8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   L2802.02.
OS   Leishmania major.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5664;
RN   [1]

```
Matches  165;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  435 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 494

Qy   61 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKQMLTESLVALQRGVFLLSDYRVT 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  495 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKQMLTESLVALQRGVFLLSDYRVT 554

Qy  121 GDTVLCYCCGLRSFRELTYQRONIPASELPVAVTSRPDCYWGRNC 166
         |||||||||||||||||||||||||||||||||||||||||||||
Db  555 GDTVLCYCCGLRSFRELTYQYQONIPASELPVAVTSRPDCYWGRNC 600
```

RESULT 5
```
Q96EP1
ID   Q96EP1;             PRELIMINARY;      PRT;   652 AA.
AC   Q96EP1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; BC012072; AAH12072.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   Pfam; PF00498; FHA; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE   652 AA;  72058 MW;  1A50773CEA679F07 CRC64;

Query Match             99.6%;  Score 939;  DB 4;  Length 652;
Best Local Similarity   99.4%;  Pred. No. 1.7e-96;  -.
Matches  165;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
         ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db  464 CPLQGSHALCTCCFQPMPDRRVEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 523

Qy   61 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKQMLTESLVALQRGVFLLSDYRVT 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  524 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKQMLTESLVALQRGVFLLSDYRVT 583

Qy  121 GDTVLCYCCGLRSFRELTYQRONIPASELPVAVTSRPDCYWGRNC 166
         |||||||||||||||||||||||||||||||||||||||||||||
Db  584 GDTVLCYCCGLRSFRELTYQYQONIPASELPVAVTSRPDCYWGRNC 629
```

RESULT 6
```
Q810L3
ID   Q810L3;             PRELIMINARY;      PRT;   664 AA.
AC   Q810L3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RIKEN cDNA 5730484M20 gene.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Limb;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC049792; AAH49792.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 2.
DR   SMART; SM00184; RING; 2.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE   664 AA;  73871 MW;  D651BE3463DEBB6 CRC64;

Query Match             96.5%;  Score 910;  DB 11;  Length 664;
Best Local Similarity   94.0%;  Pred. No. 3.1e-93;
Matches  156;  Conservative    9;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
         |||||||| ||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db  476 CPLQGSHAICTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCFGCLAP 535

Qy   61 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKQMLTESLVALQRGVFLLSDYRVT 120
         |||||||||||||||||:|||||||||||||||||:::|||||:|||||||:||||||:|
Db  536 FCELNLGDKCLDGVLNNNYESDILKNYLATRGLTWKSVLTESLLALQRGVFMLSDYRIT 595

Qy  121 GDTVLCYCCGLRSFRELTYQRONIPASELPVAVTSRPDCYWGRNC 166
         |:|||||||||||||||||||||||||||||:|||||||||||||
Db  596 GNTVLCYCCGLRSFRELTYQRONIPASELPVTVTSRPDCYWGRNC 641
```

RESULT 7
```
Q8BJZ9
ID   Q8BJZ9;             PRELIMINARY;      PRT;   663 AA.
AC   Q8BJZ9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cell cycle checkpoint protein CHFR.
GN   5730484M20RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Body;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium;
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK077629; BAC36912.1; -.
DR   MGD; MGI:2444898; 5730484M20Rik.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 2.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE   663 AA;  73855 MW;  42BD653DCB4ADF45 CRC64;
```

```
RESULT 2
Q96SL3
ID   Q96SL3      PRELIMINARY;      PRT;   652 AA.
AC   Q96SL3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ14781.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AK027687; BAB55297.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 652 AA; 72030 MW; 572P2CE6D1743D80 CRC64;

Query Match        100.0%;  Score 943;  DB 4;  Length 652;
Best Local Similarity  100.0%;  Pred. No. 6e-97;
Matches 166;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1  CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  464  CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 523

Qy   61  FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  524  FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 583

Qy  121  GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  584  GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 629


CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AF170724; AAF91084.1; -.
DR   PDB; 1LGP; 07-AUG-02.
DR   PDB; 1LGQ; 07-AUG-02.
DR   Genew; HGNC:20455; CHFR.
DR   GO; GO:0007093; P:mitotic checkpoint; TAS.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 664 AA; 73386 MW; 141A1E7FEFAE36A2 CRC64;

Query Match        100.0%;  Score 943;  DB 4;  Length 664;
Best Local Similarity  100.0%;  Pred. No. 6.1e-97;
Matches 166;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1  CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  476  CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 535

Qy   61  FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  536  FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 595

Qy  121  GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  596  GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 641


RESULT 3
Q9NRT4
ID   Q9NRT4      PRELIMINARY;      PRT;   664 AA.
AC   Q9NRT4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cell cycle checkpoint protein CHFR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20388685; PubMed=10935642;
RA   Scolnick D.M., Halazonetis T.D.;
RT   "Chfr defines a mitotic stress checkpoint that delays entry into
RT   metaphase.";
RL   Nature 406:430-435(2000).


RESULT 4
Q9NVD5
ID   Q9NVD5;      PRELIMINARY;      PRT;   623 AA.
AC   Q9NVD5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ10796.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AK001658; BAA91817.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 623 AA; 69204 MW; 45186D33DAE52711 CRC64;

Query Match        99.6%;  Score 939;  DB 4;  Length 623;
Best Local Similarity  99.4%;  Pred. No. 1.6e-96;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 14:40:47 ; Search time 24.0148 Seconds
                                   (without alignments)
                                   2180.991 Million cell updates/sec

Title: US-10-048-046-2_COPY_476_641
Perfect score: 943
Sequence: 1 CPLQGSHALCTCCFQPMPDR.........ASELPVAVTSRPDCYWGRNC 166

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SPTREMBL_25:*
               1: sp_archea:*
               2: sp_bacteria:*
               3: sp_fungi:*
               4: sp_human:*
               5: sp_invertebrate:*
               6: sp_mammal:*
               7: sp_mhc:*
               8: sp_organelle:*
               9: sp_phage:*
               10: sp_plant:*
               11: sp_rodent:*
               12: sp_virus:*
               13: sp_vertebrate:*
               14: sp_unclassified:*
               15: sp_rvirus:*
               16: sp_bacteriap:*
               17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	943	100.0	306	4	Q9NT32	Q9nt32 homo sapien
2	943	100.0	652	4	Q96SL3	Q96sl3 homo sapien
3	943	100.0	664	4	Q9NRT4	Q9nrt4 homo sapien
4	939	99.6	623	4	Q9NVD5	Q9nvd5 homo sapien
5	939	99.6	652	4	Q96EP1	Q96ep1 homo sapien
6	910	96.5	664	11	Q810L3	Q810l3 mus musculu
7	901	95.5	663	11	Q8BJZ9	Q8bjz9 mus musculu
8	137.5	14.6	473	10	Q9SX88	Q9sx88 arabidopsis
9	91.5	9.7	292	13	Q7T3F2	Q7t3f2 brachydanio
10	88.5	9.4	806	5	Q9N9P8	Q9n9p8 leishmania
11	85	9.0	504	5	Q9UPQ4	Q9upq4 homo sapien
12	83.5	8.9	2972	5	P90891	P90891 caenorhabdi
13	83	8.8	2531	5	O16004	O16004 lytechinus
14	82.5	8.7	1023	5	Q9XYD4	Q9xyd4 dictyosteli
15	81.5	8.6	523	5	Q8IOG8	Q8iog8 drosophila
16	81.5	8.6	564	5	Q8TOF5	Q8tof5 drosophila
17	81.5	8.6	585	5	Q9N675	Q9n675 drosophila
18	81.5	8.6	1056	11	Q8COP7	Q8c0p7 mus musculu
19	81	8.6	551	2	Q8L172	Q8l172 desulfitoba
20	81	8.6	551	2	Q8GJ31	Q8gj31 desulfitoba
21	81	8.6	551	2	Q8GJ27	Q8gj27 dehalobacte
22	80.5	8.5	274	12	Q7T5I7	Q7t5i7 cryptophleb
23	80.5	8.5	336	2	O69207	O69207 actinosynne
24	80	8.5	302	10	O64762	O64762 arabidopsis
25	80	8.5	530	13	Q9VTW3	Q9vtw3 drosophila
26	80	8.5	628	13	Q7ZVC1	Q7zvc1 brachydanio
27	79	8.4	1532	4	Q9P2K1	Q9p2k1 homo sapien
28	78.5	8.3	206	4	Q8WVA4	Q8wva4 homo sapien
29	78.5	8.3	493	4	Q86XQ0	Q86xq0 homo sapien
30	78	8.3	310	11	Q9CPU8	Q9cpu8 mus musculu
31	78	8.3	500	11	Q99PN4	Q99pn4 mus musculu
32	78	8.3	639	11	Q8K450	Q8k450 mus musculu
33	77.5	8.2	446	13	Q803L6	Q803l6 brachydanio
34	77.5	8.2	601	13	Q8JH43	Q8jh43 brachydanio
35	77.5	8.2	1060	3	Q8TFW3	Q8tfw3 aspergillus
36	77.5	8.2	2215	5	Q8IBW9	Q8ibw9 plasmodium
37	77	8.2	303	4	Q7Z419	Q7z419 homo sapien
38	77	8.2	1633	11	Q8CFW7	Q8cfw7 mus musculu
39	76.5	8.1	1358	5	Q960D5	Q960d5 drosophila
40	76.5	8.1	1515	13	Q9DE37	Q9de37 brachydanio
41	76.5	8.1	1700	5	Q9VVA9	Q9vva9 drosophila
42	76.5	8.1	2841	5	Q8MLU9	Q8mlu9 drosophila
43	76.5	8.1	2931	5	Q9W2C6	Q9w2c6 drosophila
44	76	8.1	120	2	Q8GHL2	Q8ghl2 serratia ma
45	76	8.1	181	10	Q9AWV5	Q9awv5 oryza sativ

ALIGNMENTS

RESULT 1
Q9NT32
ID   Q9NT32   PRELIMINARY;   PRT;   306 AA.
AC   Q9NT32;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   DKFZP434N2420.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137561; CAB70812.1; -.
DR   PIR; T46399; T46399.
KW   Hypothetical protein.
FT   NON_TER   1   1
SQ   SEQUENCE   306 AA; 34501 MW; 6B50F04601FB2939 CRC64;

Query Match   100.0%; Score 943; DB 4; Length 306;
Best Local Similarity   100.0%; Pred. No. 2.4e-97;
Matches 166; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   118 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 177

Qy   61 FCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   178 FCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 237

Qy   121 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
            |||||||||||||||||||||||||||||||||||||||||||||
Db   238 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 283

```
DR   PROSITE; PS50023; LIM_DOMAIN_2; 4.
KW   Repeat; LIM domain; Metal-binding; zinc; Zinc-finger.
FT   ZN_FING      7     31    GATA-LIKE (POTENTIAL).
FT   DOMAIN      40     92    LIM 1.
FT   DOMAIN     101    153    LIM 2.
FT   DOMAIN     162    212    LIM 3.
FT   DOMAIN     221    275    LIM 4.
SQ   SEQUENCE   279 AA;  32072 MW;  6D8CBC4B4424BFF2 CRC64;
```

```
Query Match      8.0%;  Score 75;  DB 1;  Length 279;
Best Local Similarity  23.6%;  Pred. No. 3.2;
Matches   29;  Conservative   12;  Mismatches   48;  Indels   34;  Gaps   6;
```

```
Qy    9 LCTCCFQPMPDRR-AEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELNLG 67
          :||| :|: .:|  : .:||    ||:
Db  187 VCTACKKQLSGQRFTARDEFP-------YCLTCFCDLY---AKKCAGCTNPISGLG-G 233

Qy   68 DKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCY 127
         |:. :    :.||  ::  |:
Db  234 TKYIS--FEERQWHNDCFNCKKCSLSLVGRGFLTER------------------DDILCP 273

Qy  128 CCG 130
          ||
Db  274 DCG 276
```

RESULT 14
SLI3_RAT

```
ID   SLI3_RAT       STANDARD;       PRT;       279 AA.
AC   O35115;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Skeletal muscle LIM-protein 3 (SLIM 3) (LIM-domain protein DRAL)
DE   (Four and a half LIM domains protein 2) (FHL-2).
GN   FHL2 OR SLIM3 OR DRAL.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar; TISSUE=Brain;
RA   Tanahashi H.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 4 LIM zinc-binding domains.
```

```
DR   EMBL; AB008571; BAA23357.1; -.
DR   InterPro; IPR001781; LIM.
DR   Pfam; PF00412; LIM; 4.
DR   ProDom; PD000094; LIM; 4.
DR   SMART; SM00132; LIM; 4.
DR   PROSITE; PS00478; LIM_DOMAIN_1; 4.
DR   PROSITE; PS50023; LIM_DOMAIN_2; 4.
KW   Repeat; LIM domain; Metal-binding; zinc; Zinc-finger.
FT   ZN_FING      7     31    GATA-LIKE (POTENTIAL).
FT   DOMAIN      40     92    LIM 1.
FT   DOMAIN     101    153    LIM 2.
FT   DOMAIN     162    212    LIM 3.
FT   DOMAIN     221    275    LIM 4.
SQ   SEQUENCE   279 AA;  32086 MW;  9A9DBE5935034173 CRC64;
```

```
Query Match      8.0%;  Score 75;  DB 1;  Length 279;
Best Local Similarity  23.6%;  Pred. No. 3.2;
Matches   29;  Conservative   12;  Mismatches   48;  Indels   34;  Gaps   6;
```

```
Qy    9 LCTCCFQPMPDRR-AEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELNLG 67
          :||| :|: .:|  : .:||    ||:
Db  187 VCTACKKQLSGQRFTARDEFP-------YCLTCFCDLY---AKKCAGCTNPISGLG-G 233

Qy   68 DKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCY 127
         |:. :    :.||  ::  |:
Db  234 TKYIS--FEERQWHNDCFNCKKCSLSLVGRGFLTER------------------DDILCP 273

Qy  128 CCG 130
          ||
Db  274 DCG 276
```

RESULT 15
OPCA_NOSPU

```
ID   OPCA_NOSPU       STANDARD;       PRT;       465 AA.
AC   P48971;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Putative OxPP cycle protein opcA.
GN   OPCA.
OS   Nostoc punctiforme.
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=63737;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29133 / PCC 73102;
RX   MEDLINE=95175603; PubMed=7870816;
RA   Summers M.L., Meeks J.C., Chu S., Wolf R.E. Jr.; strain ATCC 29133
RT   "Nucleotide sequence of an operon in Nostoc sp. strain ATCC 29133
RT   encoding four genes of the oxidative pentose phosphate cycle.";
RL   Plant Physiol. 107:267-268(1995).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE FUNCTIONAL ASSEMBLY OF GLUCOSE
CC       6-PHOSPHATE DEHYDROGENASE.
```

```
DR   EMBL; L32796; AAA50771.1; -.
DR   InterPro; IPR004555; OpcA.
DR   InterPro; IPR002477; PG_binding.
DR   Pfam; PF01471; PG_binding_1; 1.
DR   TIGRFAMs; TIGR00534; OpcA; 1.
SQ   SEQUENCE   465 AA;  50627 MW;  2E2D0B0D90E23931 CRC64;
```

```
Query Match      8.0%;  Score 75;  DB 1;  Length 465;
Best Local Similarity  22.8%;  Pred. No. 5.6;
Matches   38;  Conservative   19;  Mismatches   68;  Indels   42;  Gaps   7;
```

```
Qy    1 CPL--QGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCT---RTGCY 55
          ||:  :..|: |||: .::.|:: | :|  :.|:||    :|::         .|.||
Db  183 CPIQKQSSTLICCEYITLSGTAALERIGGMIPALLIGGLPKF--LWKATPDPNNGLF 240

Qy   56 GCLAPFCELNLGDKCLDGVLNNSYESDILK-NVLATRGL----------TWKNMLITES 103
         .:|..:|:.::    |:.    .:|:   | : ::||        ::
Db  241 KRLAAICNNVIVDSC----NFNEPESDLLRLRELVEAGVPLADLNWRRLASWQELTAEA 295

Qy  104 LVALQRGVFLLSDYRVTGDTVLCYCCGIRSFRELTYQRQNIPASEL 150
         :.|:|       :      |:  |          |   ::|
Db  296 YDSPKRR----------AALRLDRVTIDYEKGNPAQAL 324
```

Search completed: May 7, 2004, 14:46:34
Job time : 6.60929 secs

RESULT 12

```
ID   LPCA_YERPE     STANDARD;      PRT;    193 AA.
AC   Q8ZBY7;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phosphoheptose isomerase (EC 5.-.-.-).
GN   LPCA OR GMHA OR YPO3243 OR YO947.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / Biovar Orientalis;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KIM5 / Biovar Mediaevalis;
RX   MEDLINE=22137863; PubMed=12142430;
RA   Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA   Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA   Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA   Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA   Perry R.D.;
RT   "Genome sequence of Yersinia pestis KIM.";
RL   J. Bacteriol. 184:4601-4611(2002).
CC   -!- FUNCTION: Involved in synthesis of glyceromannoheptose 7-phosphate
CC       (By similarity).
CC   -!- PATHWAY: Inner core lipopolysaccharide biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the SIS family. LpcA subfamily.
```

```
DR   EMBL; AJ414156; CAC92478.1; -.
DR   EMBL; AE013698; AAM84528.1; -.
DR   PIR; AB0394; AB0394.
DR   HAMAP; MF_00067; -; 1.
DR   InterPro; IPR004515; GmhA.
DR   InterPro; IPR001347; SIS.
DR   Pfam; PF01380; SIS; 1.
DR   TIGRFAMs; TIGR00441; gmhA; 1.
KW   Isomerase; Lipopolysaccharide biosynthesis; Complete proteome.
FT   CONFLICT    109    109    D -> G (IN REF. 2).
SQ   SEQUENCE   193 AA;  20987 MW;  643FB0C34452D79 CRC64;
```

```
Query Match           8.0%;   Score 75;  DB 1;  Length 193;
Best Local Similarity 29.2%;  Pred. No. 2.1;
Matches   21;  Conservative  16;  Mismatches  27;  Indels  8;  Gaps  3;

Qy    83 DILKNYLATRGLTWKNMLTS--SLVALQRGVFLLSD-YRVTGDTVLC-----YCCGLRSF 134
         |:::|  |         :   :::|| |:||   :       |   :
Db     4 DLIRSELNERADTLANFLKDDSNIDIQRAAILLADSFKAGGKVLSCGNGGSHCDAMHFA 63
```

```
Qy   135 RELTYQYRQNIP 146
             ||| :||:| |
Db    64 EELTGRYRENRP 75
```

RESULT 13

```
ID   SLI3_MOUSE     STANDARD;      PRT;    279 AA.
AC   O70433; P97448;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Skeletal muscle LIM-protein 3 (SLIM 3) (LIM-domain protein DRAL)
DE   (Four and a half LIM domains protein 2) (FHL-2).
GN   FHL2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Brain;
RA   Chan K.K., Tsui S.K.W., Lee C.Y., Fung K.P., Waye M.M.Y.;
RT   "The cloning, sequencing and characterization of a mouse FHL2, which
RT   contains four and a half LIM domains.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=99160848; PubMed=10049693;
RA   Morgan M.J., Madgwick A.J.A.;
RT   "The LIM proteins FHL1 and FHL3 are expressed differently in skeletal
RT   muscle.";
RL   Biochem. Biophys. Res. Commun. 255:245-250(1999).
RN   [3]
RP   SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
RX   MEDLINE=20368180; PubMed=10906474;
RA   Chu P.-H., Ruiz-Lozano P., Zhou Q., Cai C., Chen J.;
RT   "Expression patterns of FHL/SLIM family members suggest important
RT   functional roles in skeletal muscle and cardiovascular system.";
RL   Mech. Dev. 95:259-265(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RA   Starzinski-Powitz A., Martin B., Eckerdt F.;
RT   "Isolation of the mouse homolog mDRAL from skeletal muscle derived
RT   myoblasts.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- TISSUE SPECIFICITY: Highly expressed in heart but also detectable
CC       in brain and skeletal muscle.
CC   -!- SIMILARITY: Contains 4 LIM zinc-binding domains.
```

```
DR   EMBL; AF055889; AAC12770.1; -.
DR   EMBL; U77040; AAB19211.2; -.
DR   EMBL; AF114381; AAD53230.1; -.
DR   EMBL; AF153340; AAD34170.1; -.
DR   MGD; MGI:1338762; Fhl2.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro...; IDA.
DR   InterPro; IPR001781; LIM.
DR   Pfam; PF00412; LIM; 4.
DR   ProDom; PD000094; LIM; 4.
DR   SMART; SM00132; LIM; 4.
DR   PROSITE; PS00478; LIM_DOMAIN_1; 4.
```

```
RN   [9]
RP   CARBOHYDRATE-LINKAGE SITE THR-109.
RX   MEDLINE=92184750; PubMed=1544894;
RA   Harris R.J., Ling V.T., Spellman M.W.;
RT   "O-linked fucose is present in the first epidermal growth factor
RT   domain of factor XII but not protein C.";
RL   J. Biol. Chem. 267:5102-5107(1992).
RN   [10]
RP   VARIANT WASHINGTON D.C. SER-590.
RX   MEDLINE=90046788; PubMed=2510163;
RA   Miyata T., Kawabata S.-I., Iwanaga S., Takahashi I., Alving B.,
RA   Saito H.;
RT   "Coagulation factor XII (Hageman factor) Washington D.C.: inactive
RT   factor XIIa results from Cys-571-->Ser substitution.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:8319-8322(1989).
RN   [11]
RP   VARIANT LOCARNO PRO-372.
RX   MEDLINE=94255559; PubMed=8049433;
RA   Hovinga J.K., Schaller J., Stricker H., Wuillemin W.A., Furlan M.,
RA   Laemmle B.;
RT   "Coagulation factor XII Locarno: the functional defect is caused by
RT   the amino acid substitution Arg-353-->Pro leading to loss of a
RT   kallikrein cleavage site.";
RL   Blood 84:1173-1181(1994).
RN   [12]
RP   VARIANT TENRI CYS-53.
RX   MEDLINE=99290785; PubMed=10361128;
RA   Kondo S., Tokunaga F., Kawano S., Oono Y., Kumagai S., Koide T.;
RT   "Factor XII Tenri, a novel cross-reacting material negative factor XII
RT   deficiency, occurs through a proteasome-mediated degradation.";
RL   Blood 93:4300-4308(1999).
CC   -!- FUNCTION: Factor XII is a serum glycoprotein that participates in
CC       the initiation of blood coagulation, fibrinolysis, and the
CC       generation of bradykinin and angiotensin.
CC   -!- CATALYTIC ACTIVITY: Cleaves selectively Arg-|-Ile bonds in factor
CC       VII to form factor VIIa and factor XI to form factor XIa.
CC   -!- PTM: O- AND N-GLYCOSYLATED.
CC   -!- DISEASE: Defects in F12 do not cause any clinical symptoms. The
CC       sole effect is that whole-blood clotting time is prolonged.
CC   -!- MISCELLANEOUS: Factor XII, prekallikrein, and HMW kininogen form a
CC       complex bound to an anionic surface. Prekallikrein is cleaved by
CC       factor XII to form kallikrein, which then cleaves factor XII first
CC       to alpha-factor XIIa and then to beta-factor XIIa. Alpha-factor
CC       XIIa activates factor XI to factor XIa.
CC   -!- SIMILARITY: Belongs to peptidase family S1.
CC   -!- SIMILARITY: Contains 2 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 fibronectin type I domain.
CC   -!- SIMILARITY: Contains 1 fibronectin type II domain.
CC   -!- SIMILARITY: Contains 1 kringle domain.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M31315; AAA70225.1; -.
DR   EMBL; AF538691; AAM97932.1; -.
DR   EMBL; M11723; AAA51986.1; -.
DR   EMBL; M17466; AAB59490.1; -.
DR   EMBL; M17464; AAB59490.1; JOINED.
DR   EMBL; M17465; AAB59490.1; JOINED.
DR   EMBL; M13147; AAA70224.1; -.
DR   EMBL; U71274; AAB51203.1; -.
DR   PIR; A29411; KFHU12.
DR   HSSP; P00763; 1DPO.
DR   MEROPS; S01.211; -.
DR   Genew; HGNC:3530; F12.
DR   MIM; 234000; -.
DR   GO; GO:0003805; F:blood coagulation factor XI activity; TAS.
DR   GO; GO:0003806; F:blood coagulation factor XII activity; TAS.

DR   GO; GO:0008236; F:serine-type peptidase activity; TAS.
DR   GO; GO:0007596; P:blood coagulation; TAS.
DR   InterPro; IPR009003; Cys_Ser_trypsin.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR000562; FN_Type_II.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR000001; Kringle.
DR   InterPro; IPR001254; Peptidase_S1.
DR   InterPro; IPR001314; Peptidase_S1A.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00039; fn1; 1.
DR   Pfam; PF00040; fn2; 1.
DR   Pfam; PF00051; kringle; 1.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PRINTS; PR00013; FNTYPEII.
DR   PRINTS; PR00018; KRINGLE.
DR   ProDom; PD000995; FN_Type_II; 1.
DR   ProDom; PD000395; Kringle; 1.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00058; FN1; 1.
DR   SMART; SM00059; FN2; 1.
DR   SMART; SM00130; KR; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 2.
DR   PROSITE; PS01253; FIBRONECTIN_1; 1.
DR   PROSITE; PS00023; FIBRONECTIN_2; 1.
DR   PROSITE; PS00021; KRINGLE_1; 1.
DR   PROSITE; PS50070; KRINGLE_2; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
DR   PROSITE; PS00134; TRYPSIN_HIS; 1.
DR   PROSITE; PS00135; TRYPSIN_SER; 1.
KW   Glycoprotein; Blood coagulation; Plasma; Kringle; Serine protease;
KW   Hydrolase; Fibrinolysis; Signal; EGF-like domain; Repeat; Zymogen;
KW   Polymorphism; Disease mutation.
FT   SIGNAL        1     19
FT   CHAIN        20    372       ALPHA-FACTOR XIIA HEAVY CHAIN.
FT   CHAIN       373    615       ALPHA-FACTOR XIIA LIGHT CHAIN.
FT   CHAIN       354    362       BETA-FACTOR XIIA PART 1.
FT   CHAIN       373    615       BETA-FACTOR XIIA PART 2.
FT   DOMAIN       47     88       FIBRONECTIN TYPE-II.
FT   DOMAIN       94    131       EGF-LIKE 1.
FT   DOMAIN      133    173       FIBRONECTIN TYPE-I.
FT   DOMAIN      174    210       EGF-LIKE 2.
FT   DOMAIN      217    295       KRINGLE.
FT   DOMAIN      296    349       PRO-RICH.
FT   DOMAIN      373    615       SERINE PROTEASE.
FT   CARBOHYD    109    109       O-LINKED (FUC).
FT   CARBOHYD    249    249       N-LINKED (GLCNAC...).
FT   CARBOHYD    299    299       O-LINKED (POTENTIAL).
FT   CARBOHYD    305    305       O-LINKED (POTENTIAL).
FT   CARBOHYD    308    308       O-LINKED (POTENTIAL).
FT   CARBOHYD    328    328       O-LINKED (POTENTIAL).
FT   CARBOHYD    329    329       O-LINKED (POTENTIAL).

Query Match      8.0%;  Best Local Similarity   20.7%;   Score 75.5;  DB 1;  Length 615;
Matches   29;  Conservative  13;  Mismatches  35;  Indels  63;  Gaps  6;  Pred. No. 6.9;

Qy     1 CP--LQGSHALCTCCFQPM----------------------------PDRRAEREQDPRVAP   32
           :|      ||    | :|                              |    :|       |:
Db   121 CPQHLTGNHCQKEKCFEPQLLRFFHKNEIWYRTEQAAVARCQCKGPDAHCQ-----RLAS  175

Qy    33 QQCAV-------CLQPFCHLYWGCTRTCGYCLAPFCSLNLGDKCLDGVLNNNSYESDI    84
           ||:|        ||::| |     || |||   |:|| |   |    :     | |:::
Db   176 QACRTNPCLHGGRCLEVEGHRLCHCP----VGYTGPFCDVDTKASCYDG----------  220

Qy    85 LKNYLATRGLTWKNMLTBSL   104
           |||:    :   :|
Db   221 ------RGLSYRGLARTTL   233
```

```
CC   -!- PATHWAY: Tyrosine catabolism; fifth step.
CC   -!- SUBUNIT: Homodimer.
CC   -!- TISSUE SPECIFICITY: MAINLY IN LIVER AND KIDNEY. LOWER LEVELS ARE
CC       ALSO DETECTED IN MANY OTHER TISSUES.
CC   -!- DISEASE: Defects in FAH are the cause of tyrosinemia type I
CC       [MIM:276700], an autosomal recessive inborn error of metabolism.
CC       It occurs in both an acute and a chronic form. Clinical
CC       characteristics of the acute form include hepatic failure and
CC       death in infancy, whereas children with the chronic form have
CC       renal tubular dysfunction and hypophosphatemic rickets,
CC       progressive liver disease with development of hepatocellular
CC       carcinoma. Dietary treatment with restriction of tyrosine and
CC       phenylalanine alleviates the rickets, but liver transplantation
CC       has so far been the only definite treatment. Tyrosinemia type I is
CC       a rare condition, except in the Saguenay-lac-St-Jean region
CC       (province of Quebec, Canada) where the frequency is 1/1,846
CC       newborns as the result of a founder effect.
CC   -!- SIMILARITY: Belongs to the FAH family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M55150; AAA52422.1; -.
DR   EMBL; BC002527; AAH02527.1; -.
DR   EMBL; X51728; CAA36016.1; -.
DR   PIR; A37926; A37926.
DR   Genew; HGNC:3579; FAH.
DR   GK; P16930; -.
     MIM; 276700; -.
DR   GO; GO:0004334; F:fumarylacetoacetase activity; TAS.
DR   GO; GO:0006572; P:tyrosine catabolism; TAS.
DR   InterPro; IPR002529; FAA_hydrolase.
DR   InterPro; IPR005959; Fum_ac_acetase.
DR   Pfam; PF01557; FAA_hydrolase; 1.
DR   TIGRFAMs; TIGR01266; fum_ac_acetase; 1.
KW   Hydrolase; Phenylalanine catabolism; Tyrosine catabolism;
KW   Disease mutation.
FT   VARIANT      16     16       N -> I (in tyrosinemia type I; loss of
FT                                activity).
FT                                /FTId=VAR_005205.
FT   VARIANT      62     62       F -> C (in tyrosinemia type I; loss of
FT                                activity).
FT                                /FTId=VAR_005206.
FT   VARIANT      64     64       Q -> H (in tyrosinemia type I; many
FT                                patients of Pakistani origin).
FT                                /FTId=VAR_005207.
FT   VARIANT     134    134       A -> D (in tyrosinemia type I; chronic;
FT                                loss of activity).
FT                                /FTId=VAR_005208.
FT   VARIANT     158    158       G -> D (in tyrosinemia type I).
FT                                /FTId=VAR_005209.

Query Match       8.0%;  Score 75.5;  DB 1;  Length 419;
Best Local Similarity  24.3%;  Pred. No. 4.5;
Matches   35;  Conservative  10;  Mismatches  54;  Indels  45;  Gaps   6;

Qy     17  MPDRRABREQDPRVAPQQC------------------------AVCLQPFCHLYW-----G  48
                                                     ||   |:||
Db    270  MPFAVPNPKQDPRPLPVLCHDEPYTFDINLSVNLKGEGMSQAATICKSNFXYMYWTMLQQ  329

Qy     49  CTRTGCYGCLAPFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKOMLTESLVALQ  108
               |||      :: :     |||   ||:      ||: |   ::|  |||
Db    330  LTHHSVNG-----CNLRPGDLLASGTISGPEPB-----NFGSMLELSWKGTKPIDLGNGQ  379

Qy    109  RGVFLLSDYRVTGDTVLC--YCCG  130
               ||| |::|||     |||
Db    380  TRKFLLD-----GDEVIITGYCQG  398
```

```
RESULT 11
FA12_HUMAN
ID   FA12_HUMAN     STANDARD;      PRT;    615 AA.
AC   P00748; P78339;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Coagulation factor XII precursor (EC 3.4.21.38) (Hageman factor)
DE   (HAF).
GN   F12.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88007593; PubMed=2888762;
RA   Cool D.E., McGillivray R.T.A.;
RT   "Characterization of the human blood coagulation factor XII gene.
RT   Intron/exon gene organization and analysis of the 5'-flanking
RT   region.";
RL   J. Biol. Chem. 262:13662-13673(1987).
RN   [2]
RP   SEQUENCE FROM N.A. AND VARIANTS ALA-207; ASP-545 AND HIS-605.
RA   Rieder M.J., Armel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,
RA   Rajkumar N., Toth E.J., Yi O., Nickerson D.A.;
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 4-615 FROM N.A.
RX   MEDLINE=86176794; PubMed=3754331;
RA   Tripodi M., Citarella F., Guida S., Galeffi P., Fantoni A.,
RA   Cortese R.;
RT   "cDNA sequence coding for human coagulation factor XII (Hageman).";
RL   Nucleic Acids Res. 14:3146-3146(1986).
RN   [4]
RP   SEQUENCE OF 14-615 FROM N.A.
RX   MEDLINE=86033830; PubMed=3877053;
RA   Cool D.E., Edgell C.-J.S., Louie G.V., Zoller M.J., Brayer G.D.,
RA   McGillivray R.T.A.;
RT   "Characterization of human blood coagulation factor XII cDNA.
RT   Prediction of the primary structure of factor XII and the tertiary
RT   structure of beta-factor XIIa.";
RL   J. Biol. Chem. 260:13666-13676(1985).
RN   [5]
RP   SEQUENCE OF 146-615 FROM N.A.
RX   MEDLINE=86216049; PubMed=3011063;
RA   Que B.G., Davie E.W.;
RT   "Characterization of a cDNA coding for human factor XII (Hageman
RT   factor).";
RL   Biochemistry 25:1525-1528(1986).
RN   [6]
RP   SEQUENCE OF 20-379.
RX   MEDLINE=85182674; PubMed=3886654;
RA   McMullen B.A., Fujikawa K.;
RT   "Amino acid sequence of the heavy chain of human alpha-factor XIIa
RT   (activated Hageman factor).";
RL   J. Biol. Chem. 260:5328-5341(1985).
RN   [7]
RP   SEQUENCE OF 354-362 AND 373-615.
RX   MEDLINE=83291041; PubMed=6604055;
RA   Fujikawa K., McMullen B.A.;
RT   "Amino acid sequence of human beta-factor XIIa.";
RL   J. Biol. Chem. 258:10924-10933(1983).
RN   [8]
RP   SEQUENCE OF 561-615 FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=96133302; PubMed=8528215;
RA   Schloesser M., Hofferbert S., Bartz U., Lutze G., Lammle B., Engel W.;
RT   "The novel acceptor splice site mutation 11396(G-->A) in the factor
RT   XII gene causes a truncated transcript in cross-reacting material
RT   negative patients.";
RL   Hum. Mol. Genet. 4:1235-1237(1995).
```

```
RESULT 10
FAAA_HUMAN
ID   FAAA_HUMAN     STANDARD;      PRT;   419 AA.
AC   P16930;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Fumarylacetoacetase (EC 3.7.1.2) (Fumarylacetoacetate hydrolase)
DE   (Beta-diketonase) (FAA).
GN   FAH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91150763; PubMed=1998338;
RA   Phaneuf D., Labelle Y., Berube D., Arden K., Cavenee W., Gagne R.,
RA   Tanguay R.M.;
RT   "Cloning and expression of the cDNA encoding human
RT   fumarylacetoacetate hydrolase, the enzyme deficient in hereditary
RT   tyrosinemia: assignment of the gene to chromosome 15.";
RL   Am. J. Hum. Genet. 48:525-535(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 71-419 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=90245581; PubMed=2336361;
RA   Agsteribbe E., van Faassen H., Hartog M.V., Reversma T.,
RA   Taanman J.-W., Pannekoek H., Evers R.F., Welling G.M., Berger R.;
RT   "Nucleotide sequence of cDNA encoding human fumarylacetoacetase.";
RL   Nucleic Acids Res. 18:1887-1887(1990).
RN   [4]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255958; PubMed=9101289;
RA   St Louis M., Tanguay R.M.;
RT   "Mutations in the fumarylacetoacetate hydrolase gene causing
RT   hereditary tyrosinemia type I: overview.";
RL   Hum. Mutat. 9:291-299(1997).
RN   [5]
RP   VARIANT TYROSINEMIA TYPE I ILE-106.
RX   MEDLINE=93016844; PubMed=1401055;
RA   Phaneuf D., Lambert M., Laframboise R., Mitchell G., Lettre F.,
RA   Tanguay R.M.;
RT   "Type 1 hereditary tyrosinemia. Evidence for molecular heterogeneity
RT   and identification of a causal mutation in a French Canadian
RT   patient.";
RL   J. Clin. Invest. 90:1185-1192(1992).
RN   [6]
RP   VARIANT TYROSINEMIA TYPE I ASP-134.

RX   MEDLINE=93372857; PubMed=8364576;
RA   Labelle Y., Phaneuf D., Leclerc B., Tanguay R.M.;
RT   "Characterization of the human fumarylacetoacetate hydrolase gene and
RT   identification of a missense mutation abolishing enzymatic
RT   activity.";
RL   Hum. Mol. Genet. 2:941-946(1993).
RN   [7]
RP   VARIANT TYROSINEMIA TYPE I GLY-166.
RX   MEDLINE=93306316; PubMed=8318997;
RA   Grompe M., Al-Dhalimy M.;
RT   "Mutations of the fumarylacetoacetate hydrolase gene in four patients
RT   with tyrosinemia type I.";
RL   Hum. Mutat. 2:85-93(1993).
RN   [8]
RP   VARIANT TYROSINEMIA TYPE I VAL-233.
RX   MEDLINE=95029364; PubMed=7942842;
RA   Rootwelt H., Berger R., Gray G., Kelly D.A., Coskun T.,
RA   Kvittingen E.A.;
RT   "Novel splice, missense, and nonsense mutations in the
RT   fumarylacetoacetase gene causing tyrosinemia type 1.";
RL   Am. J. Hum. Genet. 55:653-658(1994).
RN   [9]
RP   VARIANT TYROSINEMIA TYPE I TRP-341.
RX   MEDLINE=95067974; PubMed=7797370;
RA   Rootwelt H., Brodtkorb E., Kvittingen E.A.;
RT   "Identification of a frequent pseudodeficiency mutation in the
RT   fumarylacetoacetase gene, with implications for diagnosis of
RT   tyrosinemia type I.";
RL   Am. J. Hum. Genet. 55:1122-1127(1994).
RN   [10]
RP   VARIANTS TYROSINEMIA TYPE I ASP-134 AND LEU-342.
RX   MEDLINE=94274193; PubMed=8005583;
RA   Rootwelt H., Chou J., Gahl W.A., Berger R., Coskun T., Brodtkorb E.,
RA   Kvittingen E.A.;
RT   "Two missense mutations causing tyrosinemia type 1 with presence and
RT   absence of immunoreactive fumarylacetoacetase.";
RL   Hum. Genet. 93:615-619(1994).
RN   [11]
RP   VARIANTS TYROSINEMIA TYPE I SER-337 AND GLY-381.
RX   MEDLINE=95276752; PubMed=7757089;
RA   St Louis M., Poudrier J., Phaneuf D., Leclerc B., Laframboise R.,
RA   Tanguay R.M.;
RT   "Two novel mutations involved in hereditary tyrosinemia type I.";
RL   Hum. Mol. Genet. 4:319-320(1995).
RN   [12]
RP   VARIANT TYROSINEMIA TYPE I GLY-234.
RX   MEDLINE=96055520; PubMed=7550234;
RA   Hahn S.H., Krasnewich D., Brantly M., Kvittingen E.A., Gahl W.A.;
RT   "Heterozygosity for an exon 12 splicing mutation and a W234G missense
RT   mutation in an American child with chronic tyrosinemia type 1.";
RL   Hum. Mutat. 6:66-73(1995).
RN   [13]
RP   VARIANTS TYROSINEMIA TYPE I ARG-193 AND VAL-369.
RX   MEDLINE=96140750; PubMed=8557261;
RA   Ploos van Amstel J.K., Bergman A.J.I.W., van Beurden E.A.C.M.,
RA   Roijers J.F.M., Peelen T., van den Berg I.E.T., Poll-The B.T.,
RA   Kvittingen E.A., Berger R.;
RT   "Hereditary tyrosinemia type 1: novel missense, nonsense and splice
RT   consensus mutations in the human fumarylacetoacetate hydrolase gene;
RT   variability of the genotype-phenotype relationship.";
RL   Hum. Genet. 97:51-59(1996).
RN   [14]
RP   VARIANTS TYROSINEMIA TYPE I ASP-158; LEU-261; HIS-405 AND SER-366 DEL.
RX   MEDLINE=98295572; PubMed=9653815;
RA   Bergman A.J.I.W., van den Berg I.E.T., Brink W., Poll-The B.T.,
RA   Ploos van Amstel J.K., Berger R.;
RT   "Spectrum of mutations in the fumarylacetoacetate hydrolase gene of
RT   tyrosinemia type 1 patients in northwestern Europe and Mediterranean
RT   countries.";
RL   Hum. Mutat. 12:19-26(1998).
CC   -!- CATALYTIC ACTIVITY: 4-fumarylacetoacetate + H(2)O = acetoacetate +
CC       fumarate.
CC   -!- PATHWAY: Phenylalanine catabolism; sixth step.
```

RL   Nucleic Acids Res. 26:477-485(1998).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=98344072; PubMed=9677405;
RA   Gerik K.J., Li X., Pautz A., Burgers P.M.;
RT   "Characterization of the two small subunits of Saccharomyces
RT   cerevisiae DNA polymerase delta.";
RL   J. Biol. Chem. 273:19747-19755(1998).
CC   -!- FUNCTION: REQUIRED FOR REPLICATION OF THE LEADING DNA STRAND AND
CC   FOR COMPLETION OF LAGGING STRAND SYNTHESIS.
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC   + (DNA)(N).
CC   -!- SUBUNIT: HETEROTRIMER WITH SUBUNITS OF 125 KDA, 58 KDA AND 55 KDA.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Belongs to the DNA polymerase delta/II small subunit
CC   family.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; D50324; BAA08859.1; -.
DR   EMBL; X87611; CAA60928.1; -.
DR   EMBL; Z49506; CAA89528.1; -.
DR   PIR; S55194; S55194.
DR   GermOnline; 141842; -.
DR   SGD; S0003766; HYS2.
KW   Transferase; DNA-directed DNA polymerase; DNA replication;
KW   Nuclear protein.
FT   CONFLICT   156   156   L -> H (IN REF. 1).
FT   CONFLICT   465   465   S -> N (IN REF. 1).
SQ   SEQUENCE   487 AA; 55296 MW; F9B200BFB97A2C07 CRC64;

Query Match   8.1%; Score 76.5; DB 1; Length 487;
Best Local Similarity   27.3%; Pred. No. 4.2;
Matches   30; Conservative   19; Mismatches   52; Indels   9; Gaps   4;

Qy   54 CYGCLAPFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNML-TESLVALQRGVP 112
                || | : ||: |||:: |||: || : :|| ||||||:||||
Db   104 CVGSI--YCEMKYKPNVLDEVINDTYGAPDLTKSYTDKEGGSDEIMLEDESGRVLLVGDF 161

Qy   113 LLSDYRVTGDTVLCYCCGLR---SFRELTYQYRQNIPASELPVAVTSRP 158
                 : ||| |: | |: ||:   :|:: |: || : : :
Db   162 IRSTPPITG--VVVGILGMEARAGTFQVLDICYPTPLPQNPPPAPIATCP 209

RESULT 9
KDGE_HUMAN
ID   KDGE_HUMAN   STANDARD;   PRT;   567 AA.
AC   P52429; Q9UKQ3;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Diacylglycerol kinase, epsilon (EC 2.7.1.107) (Diglyceride kinase)
DE   (DGK-epsilon) (DAG kinase epsilon).
GN   DGKE OR DAGK5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Umbilical vein endothelial cells;
RX   MEDLINE=96215320; PubMed=8626589;
RA   Tang W., Bunting M., Zimmerman G.A., McIntyre T.M., Prescott S.M.;
RT   "Molecular cloning of a novel human diacylglycerol kinase highly
RT   selective for arachidonate-containing substrates.";
RL   J. Biol. Chem. 271:10237-10241(1996).
RN   [2]

RP   SEQUENCE OF 1-154 FROM N.A.
RX   MEDLINE=20035825; PubMed=10571048;
RA   Tang W., Bardien S., Bhattacharya S.S., Prescott S.M.;
RT   "Characterization of the human diacylglycerol kinase epsilon gene and
RT   its assessment as a candidate for inherited retinitis pigmentosa.";
RL   Gene 239:185-192(1999).
CC   -!- FUNCTION: Highly selective for arachidonate-containing species of
CC   diacylglycerol (DAG). May terminate signals transmitted through
CC   arachidonoyl-DAG or may contribute to the synthesis of
CC   phospholipids with defined fatty acid composition.
CC   -!- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol = ADP + 1,2-
CC   diacylglycerol 3-phosphate.
CC   -!- TISSUE SPECIFICITY: Expressed predominantly in testis.
CC   -!- SIMILARITY: Belongs to the eukaryotic diacylglycerol kinase
CC   family.
CC   -!- SIMILARITY: Contains 2 zinc-dependent phorbol-ester and DAG
CC   binding domains.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U49379; AAC50497.1; -.
DR   EMBL; AF136745; AAD45666.1; -.
DR   Genew; HGNC:2852; DGKE.
DR   MIM; 601440; -.
DR   GO; GO:0005524; F:ATP binding; TAS.
DR   GO; GO:0004143; F:diacylglycerol kinase activity; TAS.
DR   GO; GO:0008654; P:phospholipid biosynthesis; TAS.
DR   InterPro; IPR002219; DAG_PE-bind.
DR   InterPro; IPR000756; DAGKa.
DR   InterPro; IPR001206; DAGKc.
DR   Pfam; PF00130; DAG_PE-bind; 2.
DR   Pfam; PF00609; DAGKa; 1.
DR   Pfam; PF00781; DAGKc; 1.
DR   ProDom; PD002939; DAGKa; 1.
DR   ProDom; PD005043; DAGKc; 1.
DR   SMART; SM00109; C1; 2.
DR   SMART; SM00045; DAGKa; 1.
DR   SMART; SM00046; DAGKc; 1.
DR   PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
DR   PROSITE; PS00081; DAG_PE_BIND_DOM_2; 2.
DR   PROSITE; PS50081; DAG_PE_BIND; 2.
KW   Transferase; Kinase; Phorbol-ester binding; Multigene family;
KW   Transmembrane; Repeat.
FT   TRANSMEM   22   42   POTENTIAL.
FT   TRANSMEM   60   456   POTENTIAL.
FT   DOMAIN   436   456   PHORBOL-ESTER AND DAG BINDING 1.
FT   DOMAIN   108   177   PHORBOL-ESTER AND DAG BINDING 2.
FT   DOMAIN   125   350   CATALYTIC-A (POTENTIAL).
FT   DOMAIN   217   524   CATALYTIC-B (POTENTIAL).
FT   DOMAIN   369   .
SQ   SEQUENCE   567 AA; 63927 MW; BC334AD15FB4D0B4 CRC64;

Query Match   8.1%; Score 76; DB 1; Length 567;
Best Local Similarity   21.6%; Pred. No. 5.6;
Matches   36; Conservative   54; Mismatches   21; Indels   56; Gaps   10;

Qy   3 LQGSHALCTCCFQPMPD---RRAER------------EQDPRVAPQQ---------CAVCL 39
                || ||| | :| |||| : | : |:
Db   82 LQG--AFCDCCGLRVDEGCLRKADKRFQCKEIMLKNDTKVLDAMPHHWIRGNVPLCSYCM 139

Qy   40 QPFCHLYWGCTRTGC-YGCLAPFCELNLGDKCLDGVLNNNSYESDILKN------YLAT- 91
                :| | | | | |: |: : | |:: |:||:| : ||:| || |:|||
Db   140 --VCKQQCGCQPKLCDYRCI--WCQKTVHDECMKNSLKNEKCDFGEFKNLIIPPSVLTSI 195

Qy   92 ------------RGLTWKNM--LTESLVALQRGVFLLSDYRV 119
                                  ::|| | | |||: |::|: | |:||::
Db   196 NQMRKDKKTDYBVLASKLGKQWTPLIILANSRSGTNMGEGLLGEFRI 242

```
Qy        129 CG 130
              ||
Db        275 CG 276

RESULT 7
ID   TF1B_MOUSE    STANDARD;    PRT;   834 AA.
AC   Q62318; P70391;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transcription intermediary factor 1-beta (TIF1-beta) (Tripartite motif
DE   protein 28) (KRAB-A interacting protein) (KRIP-1).
GN   TRIM28 OR TIF1B OR KRIP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97133299; PubMed=8978696;
RA   le Douarin B., Nielsen A.L., Garnier J.M., Ichinose H., Jeanmougin F.,
RA   Losson R., Chambon P.;
RT   "A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic
RT   control of transcription by nuclear receptors.";
RL   EMBO J. 15:6701-6715(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=97140325; PubMed=8988806;
RA   Kim S.-S., Chen Y.-M., O'Leary E., Witzgall R., Vidal M.,
RA   Bonventre J.V.;
RT   "A novel member of the RING finger family, KRIP-1, associates with
RT   the KRAB-A transcriptional repressor domain of zinc finger
RT   proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:15299-15304(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20400347; PubMed=10940561;
RA   Cammas F., Garnier J.-M., Chambon P., Losson R.;
RT   "Correlation of the exon/intron organization to the conserved domains
RT   of the mouse transcriptional corepressor TIF1beta.";
RL   Gene 253:231-235(2000).
CC   -!- FUNCTION: Forms a complex with a KRAB-domain transcription factor
CC       and increases the efficiency of KRAB-mediated repression by
CC       recruiting SETDB1 to histone H3 (By similarity).
CC   -!- SUBUNIT: Interacts with SETDB1 and CBX3 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- SIMILARITY: Contains 2 B box-type zinc fingers.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X99644; CAA67963.1; -.
DR   EMBL; U67303; AAB17272.1; -.
DR   EMBL; AF230878; AAG02638.1; -.
DR   MGD; MGI:109274; Trim28.
DR   GO; GO:0000785; C:chromatin; IDA.
DR   InterPro; IPR003649; Bbox_C.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR000315; Znf_Bbox.
DR   InterPro; IPR001965; Znf_PHD.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00628; PHD; 1.
DR   Pfam; PF00643; zf-B_box; 2.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00502; BBC; 1.
DR   SMART; SM00336; BBOX; 2.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00249; PHD; 2.
DR   SMART; SM00184; RING; 2.
DR   PROSITE; PS50119; ZF_BBOX; 2.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 1.
DR   PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Transcription regulation; Repressor; Nuclear protein; Zinc-finger;
KW   Repeat.
FT   DOMAIN       42     49       POLY-ALA.
FT   DOMAIN       54     57       POLY-GLY.
FT   ZN_FING      66     93       RING-TYPE.
FT   ZN_FING     149    196       B BOX-TYPE 1.
FT   ZN_FING     205    246       B BOX-TYPE 2.
FT   DOMAIN      526    531       POLY-ALA.
FT   ZN_FING     625    672       PHD-TYPE.
FT   CONFLICT    530    530       A -> S (IN REF. 2).
SQ   SEQUENCE    834 AA;   88847 MW;  DE87AAA5DC67BB8B CRC64;

Query Match             8.3%;  Score 78;  DB 1;  Length 834;
Best Local Similarity  19.9%;  Pred. No. 5.4;
Matches   40;  Conservative   18;  Mismatches   59;  Indels   84;  Gaps    8;

Qy     10  CTCCFQPMDRRAEREQDPRVAPQQCAVCLQPFCHLWGCTRGCYGLAPFCELNLGD-  68
               :: : ||||||: ::||:::|| ||:: :||  ||      ||      ||-
Db     66  CGVRE------RLRPERDPRLLP-----CLHSAC--------SACLGPATPAAANNSGDG 107

Qy     69  -KCLDG------VLNNNSYESDILKNYL----------------------ATRG  93
            ||| :      :::|::|: |:|  : |::|
Db    108  GSAGDGAMVDCPVCKQQCVSKDIVENYPMRDSGSKASSDSQDANQCCTSCEDNAPATSYC 167

Qy     94  LTWKNMLIESLVALQRGVFLLSDY--RVTG------------------DT 123
            :  :  :   ::|    |:|:  ||  ||:                   ||
Db    168  VECSEPLCETCVEAHQRVKYTKDHTVRSTGPAKTRDGERTVYCNVHKHEPLVLFCESCDT 227

Qy    124  VLCYCCGLRSFRELTYQYRQN 144
            :  :  :    ||:: :::
Db    228  LTCRDCQLNAHKDHQYQFLED 248

RESULT 8
ID   DPD2_YEAST    STANDARD;    PRT;   487 AA.
AC   P46957;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   DNA polymerase delta small subunit (EC 2.7.7.7).
GN   POL31 OR HYS2 OR HUS2 OR SDP5 OR YJR006W OR J1427 OR YJR83.7.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sugimoto K., Sakamoto Y., Matsumoto K.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RA   de Haan M., Smits P.H.M., Grivell L.A.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX   MEDLINE=98083176; PubMed=9421503;
RA   Hashimoto K., Nakashima N., Ohara T., Maki S., Sugino A.;
RT   "The second subunit of DNA polymerase III (delta) is encoded by the
RT   HYS2 gene in Saccharomyces cerevisiae.";
```

```
DR   EMBL; AE016842; AAO70124.1; -.
DR   StyGene; SG?????; 1pcA.
DR   HAMAP; MF_00067; -; 1.
DR   InterPro; IPR004515; GmhA.
DR   InterPro; IPR001347; SIS.
DR   Pfam; PF01380; SIS; 1.
DR   TIGRFAMs; TIGR00441; gmhA; 1.
KW   Isomerase; Lipopolysaccharide biosynthesis; Complete proteome.
SQ   SEQUENCE   192 AA;  20896 MW;  0ABBFCCBCD6786A4 CRC64;

Query Match             8.5%;  Score 80;  DB 1;  Length 192;
Best Local Similarity  30.6%;  Pred. No. 0.66;
Matches    22;  Conservative  15;  Mismatches  27;  Indels  8;  Gaps  3;

Qy       83 DILKNYLATRGLTWKNMLTE--SLVALQRGVFLLSD-YRVTGDTVLC-----YCCGLRSF 134
               |:::||      |    |   :: ||||       |::|   : :       | :|:|
Db        4 DLIRNELNEAAETLANFLKDDANIHAIQRAAVLLADSFKAGGKVLSCGNGGSHCDAMHFA 63

Qy      135 RELTYQYRQNIP 146
               ||:|:|| |
Db       64 EELTGRYRENRP 75

RESULT 6
SLI3_HUMAN
ID   SLI3_HUMAN     STANDARD;      PRT;   279 AA.
AC   Q14192; Q13229; Q13644; Q9P294;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Skeletal muscle LIM-protein 3 (SLIM 3) (LIM-domain protein DRAL)
DE   (Four and a half LIM domains protein 2) (FHL-2).
GN   FHL2 OR SLIM3 OR DRAL.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=97294674; PubMed=9150430;
RA   Genini M., Schwalbe P., Scholl F.A., Remppis A., Mattei M.-G.,
RA   Schaefer B.W.;
RT   "Subtractive cloning and characterization of DRAL, a novel LIM-domain
RT   protein down-regulated in rhabdomyosarcoma.";
RL   DNA Cell Biol. 16:433-442(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Heart;
RX   MEDLINE=98248917; PubMed=9573400;
RA   Chan K.K., Tsui S.K.W., Lee S.M.Y., Luk S.C.W., Liew C.C., Fung K.P.,
RA   Waye M.M.Y., Lee C.Y.;
RT   "Molecular cloning and characterization of FHL2, a novel LIM domain
RT   protein preferentially expressed in human heart.";
RL   Gene 210:345-350(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leukocyte;
RX   MEDLINE=20458893; PubMed=11001931;
RA   Tanahashi H., Tabira T.;
RT   "Alzheimer's disease-associated presenilin 2 interacts with DRAL, an
RT   LIM-domain protein.";
RL   Hum. Mol. Genet. 9:2281-2289(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SEQUENCE OF 127-279 FROM N.A.
RC   TISSUE=Heart muscle;
RX   MEDLINE=96354835; PubMed=8753811;
RA   Morgan M.J., Madgwick A.J.A.;
RT   "Slim defines a novel family of LIM-proteins expressed in skeletal
RT   muscle.";
RL   Biochem. Biophys. Res. Commun. 225:632-638(1996).
CC   -!- TISSUE SPECIFICITY: Expressed only in skeletal muscle.
CC   -!- SIMILARITY: Contains 4 LIM zinc-binding domains.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L42176; AAA85333.1; -.
DR   EMBL; U29332; AAC52073.1; -.
DR   EMBL; AB038794; BAA92253.1; -.
DR   EMBL; AB038991; BAA92253.1; JOINED.
DR   EMBL; AB038992; BAA92253.1; JOINED.
DR   EMBL; AB038793; BAA92253.1; JOINED.
DR   EMBL; BC014397; AAH14397.1; -.
DR   EMBL; U60117; AAC50794.1; -.
DR   PIR; JC6565; JC6565.
DR   Genew; HGNC:3703; FHL2.
DR   MIM; 602633; -.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   InterPro; IPR001781; LIM.
DR   Pfam; PF00412; LIM; 4.
DR   ProDom; PD000094; LIM; 4.
DR   SMART; SM00132; LIM; 4.
DR   PROSITE; PS00478; LIM_DOMAIN_1; 4.
DR   PROSITE; PS50023; LIM_DOMAIN_2; 4.
KW   Repeat; LIM domain; Metal-binding; Zinc; Zinc-finger.
FT   ZN_FING       7     31       GATA-LIKE (POTENTIAL).
FT   DOMAIN       40     92       LIM 1.
FT   DOMAIN      101    153       LIM 2.
FT   DOMAIN      162    212       LIM 3.
FT   DOMAIN      221    275       LIM 4.
FT   CONFLICT    167    167       M -> G (IN REF. 1).
FT   CONFLICT    167    167       M -> K (IN REF. 3 AND 4).
SQ   SEQUENCE   279 AA;  32196 MW;  DBDC77C9BD6C4BBA CRC64;

Query Match             8.4%;  Score 79;  DB 1;  Length 279;
Best Local Similarity  23.8%;  Pred. No. 1.3;
Matches    29;  Conservative  10;  Mismatches  51;  Indels  32;  Gaps  5;

Qy        9 LCTCCFQPMDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRPGCYGCLAPFCELNLGD 68
              :||:||  |              |||:|         |  :| || |  ||||
Db      187 VCTACRKQLSGQRFTARDD-----FAYCLNCFCDLY---AKKCAGTNPISGLG-GT 234

Qy       69 KCLDGVLNNNSYESDILKNVLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCYC 128
              |:|| |:|    |:|  ||: |  |||  |  |:|  ||| :|| ||:| :|  |:: ||
Db      235 KYIS--FEERQWHNDCFNCKKCSLSLVGRGFLTER-----------------DDILCPD 274
```

```
Db    4 DLIRNELNEAAETLANFLKDDANIHAIQRAAVLLADSFKAGGKVLSCGNGGSHCDAMHFA 63

Qy  135 RELTYQYRQNIP 146
        ||| .:||:||.|
Db   64 EELTGRYRENRP 75
```

RESULT 5
LPCA_SALTY
ID   LPCA_SALTY   STANDARD;   PRT;   192 AA.
AC   Q8XGZ8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phosphoheptose isomerase (EC 5.-.-.-).
GN   LPCA OR GMHA OR STM0310 OR STY0355 OR T2540.
OS   Salmonella typhimurium, and
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Salmonella.
OX   NCBI_TaxID=602, 601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.typhi; STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;
RX   MEDLINE=22531367; PubMed=12644504;
RA   Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA   Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT   "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT   and CT18.";
RL   J. Bacteriol. 185:2330-2337(2003).
CC   -!- FUNCTION: Involved in synthesis of glyceromannoheptose 7-phosphate
CC   (By similarity).
CC   -!- PATHWAY: Inner core lipopolysaccharide biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the SIS family. LpcA subfamily.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AE008709; AAL19267.1; -.
DR   EMBL; AL627266; CAD08780.1; -.

RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
RN   [9]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX   MEDLINE=22272406; PubMed=12384590;
RA   Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA   Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA   Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA   Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA   Yu J.;
RT   "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT   through comparison with genomes of Escherichia coli K12 and O157.";
RL   Nucleic Acids Res. 30:4432-4441(2002).
RN   [10]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX   MEDLINE=22590274; PubMed=12704152;
RA   Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA   Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA   Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA   Schwartz D.C., Blattner F.R.;
RT   "Complete genome sequence and comparative genomics of Shigella
RT   flexneri serotype 2a strain 2457T.";
RL   Infect. Immun. 71:2775-2786(2003).
CC   -!- FUNCTION: Involved in synthesis of glyceromannoheptose 7-
CC   phosphate.
CC   -!- PATHWAY: Inner core lipopolysaccharide biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the SIS family. LpcA subfamily.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U32590; AAC43630.1; -.
DR   EMBL; D38582; BAA07584.1; -.
DR   EMBL; AE000131; AAC73326.1; -.
DR   EMBL; D83536; BAA77892.1; -.
DR   EMBL; U70214; AAB08644.1; ALT_INIT.
DR   EMBL; AE016756; AAN78853.1; ALT_INIT.
DR   EMBL; AE005200; AAG54547.1; -.
DR   EMBL; AP002551; BAB33672.1; -.
DR   EMBL; AE015062; AAN41932.1; ALT_INIT.
DR   EMBL; AE016979; AAP15818.1; -.
DR   PIR; A90660.
DR   PIR; A90660; G64746.
DR   PIR; G64746; G85510.
DR   PIR; G85510; G85510.
DR   EcoGene; EG13146; lpcA.
DR   HAMAP; MF_00067; -; 1.
DR   InterPro; IPR004515; GmhA.
DR   InterPro; IPR001347; SIS.
DR   Pfam; PF01380; SIS; 1.
DR   TIGRFAMs; TIGR00441; gmhA; 1.
KW   Isomerase; Lipopolysaccharide biosynthesis; Complete proteome.
SQ   SEQUENCE   192 AA; 20815 MW; 7A2C05E107910BB4 CRC64;
```

```
Query Match     8.5%;  Score 80;  DB 1;  Length 192;
Best Local Similarity  30.6%;  Pred. No. 0.66;  Matches  22;  Conservative  15;  Mismatches  27;  Indels  8;  Gaps  3;

Qy  83 DILKNYLATRGLTWKNMLTE-SLVALQRGVFLLSD-YRVTGDTVLC----YCCGLRSF 134
        |::::|| |     ||| :: | ::|||| ::|||| ||: ::  :|    ::| | :
Db   1 DILKNYLATRGLTWKNMLTE-SLVALQRGVFLLSD-YRVTGDTVLC----YCCGLRSF 134
```

RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUB=Pancreas, and Thymus;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
CC   -!- SUBCELLULAR LOCATION: Type III membrane protein (Potential).
CC   -!- INDUCTION: By interferons.
CC   -!- SIMILARITY: Belongs to the TMEM7 family.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AJ251364; CAC13976.1; -.
DR   EMBL; AK007477; BAB25057.1; -.
DR   EMBL; AK018021; BAB31041.1; -.
DR   MGD; MGI:1915025; 5830458K16Rik.
KW   Transmembrane.
FT   TRANSMEM   228   248
FT   CONFLICT   185   185   K -> N (IN REF. 2; BAB31041).
FT   CONFLICT   209   209   F -> L (IN REF. 2; BAB31041).
SQ   SEQUENCE   249 AA; 28391 MW; 2C79B36ED6F042D4 CRC64;

Query Match   8.6%;
Best Local Similarity 22.2%;   POTENTIAL.
Matches 37;   Conservative 21;   Mismatches 71;   Indels 38;   Gaps   6;

Qy   17 MPDRRA---BREQDPRVAPQQCAVCLQPF-----------CHLY------------WGCTR 51
             :||| : | :|:| :|:| ||:|
Db   36 VPDGAALGWRQHQQTVLGRFQCSRCCRSWTSAQVMILCHMPDTLKSQGQARMRIFGQKC 95

Qy   52 TGCYGCLAPFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKQMLTESLVALQRGV 111
             :||| : |||:   :   ::| :   :| : | :   ::||
Db   96 QKCFGC--QFETPKFSTELIIKRILNN------LVNYILQRYYGHRKIALTSNASLGEKV 146

Qy   112 FLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRP 158
               :   :|:
Db   147 TLDGPH---DTRNCEACSLNSHGRCALAHKVKPPRSPSPLPKSSSP 189

RESULT 4
LPCA_ECOLI
ID   LPCA_ECOLI   STANDARD;   PRT;   192 AA.
AC   P51001;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phosphoheptose isomerase (EC 5.-.-.-).
GN   LPCA OR GMHA OR TFRA OR B0222 OR C0372 OR Z0280 OR ECS0249 OR SF0272
GN   OR S0293.
OS   Escherichia coli,

OS   Escherichia coli O6,
OS   Escherichia coli O157:H7, and
OS   Shigella flexneri.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, 217992, 83334, 623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=K12 / W3110;
RX   MEDLINE=96216440; PubMed=8631969;
RA   Brooke J.S., Valvano M.A.;
RT   "Biosynthesis of inner core lipopolysaccharide in enteric bacteria
RT   identification and characterization of a conserved phosphoheptose
RT   isomerase.";
RL   J. Biol. Chem. 271:3608-3614(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=K12;
RA   Ohmori H.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=K12 / W3110;
RA   Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA   Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA   Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;
RT   "Systematic sequencing of the Escherichia coli genome: analysis of the
RT   4.0 - 6.0 min (189,987 - 281,416bp) region.";
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli;
RA   Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA   Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA   Davis R.W.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN   [7]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [8]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;

CC
DR   EMBL; AF025441; AAC39561.1; ALT_INIT.
DR   EMBL; BC015050; AAH15050.1; -.
DR   MIM; 606020; -.
DR   GO; GO:0005515; F:protein binding; TAS.
DR   GO; GO:0007154; P:cell communication; NAS.
SQ   SEQUENCE   229 AA; 24691 MW; 0EBD4006193A3106 CRC64;

Query Match   9.0%; Score 85; DB 1; Length 229;
Best Local Similarity 26.6%; Pred. No. 0.25;
Matches   38; Conservative   15; Mismatches   52; Indels   38; Gaps   6;

Qy   26 QDPRVAPQQCAVC--LQPPCHLWGCTRTGCYGCLAPPCELNiGDKCLDGVLNNNSYES 82
                ||| | |: | :||| |: |
Db   71 QPERCAVFQCAQCHAVLADSVHLAWDLSR--------SLGAVVPSRVTNNVVLEA 117

Qy   83 DILKNYLAT-RGLTWKNML----------TESLVALQRGVFLLSDYRVTGDTVLC 126
                ||:|:| :|| || ||| |: :|
Db   118 PFLVGIBGSLKGSTY-NLLFCGSCGIPVGFHLYSTHAALAALRGHFCLS-----SDKWVC 171

Qy   127 YCCGLRSFRELTYQYRQNIPASE 149
                :| : | ||| :| |||
Db   172 YLLKTKAIVNASEMDIQNVPLSE 194

RESULT 2
Z451_MOUSE
ID   Z451_MOUSE   STANDARD;   PRT; 1056 AA.
AC   Q8CO7;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Zinc finger protein 451.
GN   ZNF451.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=22354683; PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Magiott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang I.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
CC   -!- FUNCTION: May be involved in transcriptional regulation.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).

CC   -!- SIMILARITY: BELONGS TO THE KRUPPEL FAMILY OF C2H2-TYPE ZINC-
CC   FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 12 C2H2-type zinc fingers.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; AK030088; BAC26778.1; -.
DR   MGD; MGI:2138298; AI596398.
DR   InterPro; IPR007087; ZnF_C2H2.
DR   Pfam; PF00096; zf-C2H2; 7.
DR   SMART; SM00355; ZnF_C2H2; 12.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   ZN_FING   169   195   C2H2-TYPE 1.
FT   ZN_FING   212   234   C2H2-TYPE 2   (DEGENERATE).
FT   ZN_FING   253   277   C2H2-TYPE 3.
FT   ZN_FING   315   338   C2H2-TYPE 4   (ATYPICAL).
FT   ZN_FING   362   385   C2H2-TYPE 5.
FT   ZN_FING   494   517   C2H2-TYPE 6.
FT   ZN_FING   527   550   C2H2-TYPE 7.
FT   ZN_FING   604   629   C2H2-TYPE 8   (ATYPICAL).
FT   ZN_FING   634   657   C2H2-TYPE 9.
FT   ZN_FING   665   688   C2H2-TYPE 10.
FT   ZN_FING   751   774   C2H2-TYPE 11.
FT   ZN_FING   787   810   C2H2-TYPE 12.
SQ   SEQUENCE   1056 AA; 120069 MW; 7D8FBC0B50ECA622 CRC64;

Query Match   8.6%; Score 81.5; DB 1; Length 1056;
Best Local Similarity 25.0%; Pred. No. 3.1;
Matches   38; Conservative   17; Mismatches   52; Indels   45; Gaps   8;

Qy   35 CAVCLQPFC-------HLYWGCTRTG---------CYGCLAPFCELNLGDKCLD 72
                |||| | |: || ::|
Db   214 CAVCYEHFVTQQQYKDHL---LSRTAAADGHSNSLLPQIIQCYACPQCFLLFSTKDECLK 270

Qy   73 GVLNNNSYESD--ILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYVRVTGDTVLCYCC- 129
                 :: |: | :|::| | | | | |:
Db   271 HMSTKNHFHQSFKLSDNKGTARPISFPSFAKKRLVSLCKDVPF-------QVKCVACH 321

Qy   130 -GLRSFRELTYQYR---QNIPASELPVAVTSR 157
                  | :| |::| | |: |
Db   322 QTLRSHMELTAHFRVRCQNAG---PVAIAEK 349

RESULT 3
IF28_MOUSE
ID   IF28_MOUSE   STANDARD;   PRT;   249 AA.
AC   Q9ER80; Q9DD6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   28_kDa interferon responsive protein.
GN   IFRG28.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DBA/2;
RA   Meritet J.F., Dron M., Tovey M.;
RT   "Characterization of ifrg15 and ifrg28, two newly identified
RT   interferon responsive gene.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]

OM protein - protein search, using sw model

```
            GenCore version 5.1.6
    Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:     May 7, 2004, 14:37:32 ; Search time 5.60929 Seconds
                                    (without alignments)
                        1540.951 Million cell updates/sec

Title:      US-10-048-046-2 COPY 476_641
Perfect score: 943
Sequence:   1 CPLQGSHALCTCCFQPMPDR.........ASELPVAVTSRPDCYWGRNC 166

Scoring table: BLOSUM62
            Gapop 10.0 , Gapext 0.5

Searched:   141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:    141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
            Maximum Match 100%
            Listing first 45 summaries

Database :  SwissProt_42:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	85	9.0	229	1	OIP5_HUMAN	O43482 homo sapien
2	81.5	8.6	1056	1	Z451_MOUSE	Q8c0p7 mus musculu
3	81	8.6	249	1	IF28_MOUSE	Q9er80 mus musculu
4	80	8.5	192	1	LPCA_ECOLI	P51001 escherichia
5	80	8.5	192	1	LPCA_SALTY	Q8xgz8 salmonella
6	79	8.4	279	1	SLI3_HUMAN	Q14192 homo sapien
7	78	8.3	834	1	TF1B_MOUSE	Q62318 mus musculu
8	76.5	8.1	487	1	DPD2_YEAST	P46957 saccharomyc
9	76	8.1	567	1	KDGE_HUMAN	P52429 homo sapien
10	75.5	8.0	419	1	FAAA_HUMAN	P16930 homo sapien
11	75.5	8.0	615	1	FA12_HUMAN	P00748 homo sapien
12	75	8.0	193	1	LPCA_YERPE	Q8zby7 yersinia pe
13	75	8.0	279	1	SLI3_MOUSE	Q70433 mus musculu
14	75	8.0	279	1	SLI3_RAT	O35115 rattus norv
15	75	8.0	465	1	OPCA_NOSPU	P48971 nostoc punc
16	75	8.0	1180	1	RNT1_DROME	Q9vys3 drosophila
17	75	8.0	1458	1	PHLX_RABIT	Q05017 oryctolagus
18	74.5	7.9	603	1	FA12_CAVPO	Q04962 cavia porce
19	73.5	7.8	419	1	FAAA_MOUSE	P35505 mus musculu
20	73.5	7.8	419	1	FAAA_RAT	P25093 rattus norv
21	73	7.7	603	1	PGH2_CHICK	P27607 gallus gall
22	72.5	7.7	835	1	TF1B_HUMAN	Q13263 homo sapien
23	72	7.6	193	1	LPCA_PHOLL	Q7n7f7 photorhabdu
24	72	7.6	3075	1	LMA1_HUMAN	P25391 homo sapien
25	71.5	7.6	854	1	NAL6_RAT	Q63035 rattus norv
26	71	7.5	450	1	NRI1_HUMAN	Q9bu40 homo sapien
27	71	7.5	485	1	RN14_MOUSE	Q9j190 mus musculu
28	71	7.5	3718	1	LMA5_MOUSE	Q61001 mus musculu
29	70.5	7.5	192	1	YD42_AQUAE	O67344 aquifex aeo
30	70.5	7.5	351	1	Y876_METJA	Q58286 methanococc
31	70.5	7.5	387	1	SGCA_MOUSE	P82350 mus musculu
32	70.5	7.5	1135	1	VGLM_TSWV1	P36291 tomato spot
33	70	7.4	1093	1	RNT1_NEUCR	Q9nehi neurospora
34	69.5	7.4	63	1	KURT_PARTR	P58910 parabuthus
35	69.5	7.4	2139	1	CRB_DROME	P10040 drosophila
36	69	7.3	305	1	FX34_HUMAN	Q9nwn3 homo sapien
37	68.5	7.3	245	1	YPC2_ECOLI	P18128 escherichia
38	68.5	7.3	456	1	NRLI_CHICK	Q902d5 gallus gall
39	68.5	7.3	491	1	2A5R_MOUSE	Q92176 mus musculu
40	68.5	7.3	780	1	OSTA_YERPE	Q8zik3 yersinia pe
41	68.5	7.3	1122	1	TERT_MOUSE	O70372 mus musculu
42	68.5	7.3	1341	1	YL78_YEAST	Q05854 saccharomyc
43	68.5	7.3	2715	1	G156_PARPR	P13837 paramecium
44	68	7.2	342	1	ARGC_CAMJE	Q9pis0 campylobact
45	68	7.2	550	1	SYM_CHLTR	O84035 chlamydia t

ALIGNMENTS

```
RESULT 1
OIP5_HUMAN
ID   OIP5_HUMAN     STANDARD;      PRT;    229 AA.
AC   O43482; Q96BX7;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Opa-interacting protein 5.
GN   OIP5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98125741; PubMed=9466265;
RA   Williams J.M., Chen G.-C., Zhu L., Rest R.F.;
RT   "using the yeast two-hybrid system to identify human epithelial cell
RT   proteins that bind gonococcal Opa proteins: intracellular gonococci
RT   bind pyruvate kinase via their Opa proteins and require host pyruvate
RT   for growth.";
RL   Mol. Microbiol. 27:171-186(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H. Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBUNIT: Binds outer membrane protein OpaP from Neisseria
CC       gonorrhoeae.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
A;Molecule type: DNA
A;Residues: 1-1101 <GEI>
A;Cross-references: EMBL:U39644; NID:g1049339; PID:g1049343; PIDN:AAA80360.1; CESP:T10E1
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:T10E10.4
A;Introns: 93/2; 152/2; 191/3; 209/2; 283/3; 303/1; 399/3; 421/1; 440/1; 465/1; 547/3; 7
```

```
Query Match           8.0%;  Score 75.5;  DB 2;  Length 1101;
Best Local Similarity   20.3%;  Pred. No. 54;
Matches    35;  Conservative    27;  Mismatches    55;  Indels    55;  Gaps    9;
```

```
Qy     1 CPLQGSHALCTCFQPMPDRRABREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
           :   :|   ||||:|: :  :  ::|||  :  ||| |  : :|  |
Db   558 CGVGSCYSKTTCCQEPI--KLPVCSNGMISQKRCLVAAE--CGLNLECSNGGC--CPIP 610

Qy    61 FC------ELN---LGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVA 106
          ||       ::|      |::|:           :  :|:|    |:|:: :|:|
Db   611 FCPNGVTARGRCSQVNGCPMGQACMEGLC------CPLPK---CSNGITSLGICTRTLDC 661

Qy   107 LQRGV----------FLLSDYRVTGDTVLCY----------CCGLRS 133
          :||            :    :|:|               ||||
Db   662 GRIGVDCNNGACCPLPTCPNNIASSQRCSGGCTNCCPVGQTCMNGGCCDLPS 713
```

```
Search completed: May  7, 2004, 14:50:10
Job time : 11.5404 secs
```

```
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: H84724
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss. D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: H84724
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-543 <STO>
A;Cross-references: GB:AB002093; NID:g4887759; PIDN:AAD32295.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g31770
A;Map position: 2

Query Match         8.0%;     Score 75.5;  DB 2;  Length 543;
Best Local Similarity  25.3%;    Pred. No. 27;
Matches   24;  Conservative  10;  Mismatches  32;  Indels  29;  Gaps   5;

Qy    10 CTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYW-----------GCRTGCYGC 57
                                                      CGHPYCKTCWAGYITTKIEDGPGCLRV--KC 173
Db   128 CGICFESY-----TREBIARVS

Qy    58 LAPFCELNLGDKCLDGVLN--NNSYESDIIKNYL 89

Db   174 PEPSCSSAAVGKDMIEDVTETKVNEKYSRYILRSYV 208

RESULT 14
KFHU12
coagulation factor XIIa (EC 3.4.21.38) precursor [validated] - human
N;Alternate names: Hageman factor (activated)
C;Species: Homo sapiens (man)
C;Date: 27-Nov-1985 #sequence_revision 30-Jun-1991 #text_change 08-Dec-2000
C;Accession: A29411; A26814; A00930; A25191; A22248; A21037
R;Cool, D.E.; MacGillivray, R.T.A.
J. Biol. Chem. 262, 13662-13673, 1987
A;Title: Characterization of the human blood coagulation factor XII gene. Intron/exon ge
A;Reference number: A29411; MUID:88007593; PMID:2888762
A;Accession: A29411
A;Molecule type: DNA
A;Residues: 1-615 <COO>
A;Cross-references: GB:M17466; GB:J02807; NID:g180355; PIDN:AAB59490.1; PID:g180357
R;Tripodi, M.; Citarella, F.; Guida, S.; Galeffi, P.; Fantoni, A.; Cortese, R.
Nucleic Acids Res. 14, 3146, 1986
A;Title: cDNA sequence coding for human coagulation factor XII (Hageman).
A;Reference number: A26814; MUID:86176794; PMID:3754331
A;Accession: A26814
A;Molecule type: mRNA
A;Residues: 4-615 <TRI>
A;Cross-references: GB:M31315; NID:g182291; PIDN:AAA70225.1; PID:g182292
R;Cool, D.E.; Edgell, C.J.S.; Louie, G.V.; Zoller, M.J.; Brayer, G.D.; MacGillivray, R.T
J. Biol. Chem. 260, 13666-13676, 1985
A;Title: Characterization of human blood coagulation factor XII cDNA. Prediction of the
A;Reference number: A00930; MUID:86033830; PMID:3877053
A;Accession: A00930
A;Molecule type: mRNA
A;Residues: 14-332,'S',334-615 <CO2>
A;Cross-references: GB:M11723; NID:g180358; PIDN:AAA51986.1; PID:g180359
R;Que, B.G.; Davie, E.W.
Biochemistry 25, 1525-1528, 1986
A;Title: Characterization of a cDNA coding for human factor XII (Hageman factor).
A;Reference number: A25191; MUID:86216049; PMID:3011063
A;Accession: A25191
A;Molecule type: mRNA
A;Residues: 146-378,'G',380-615 <QUE>
A;Cross-references: GB:M13147; NID:g180360; PIDN:AAA70224.1; PID:g180361
R;McMullen, B.A.; Fujikawa, K.
J. Biol. Chem. 260, 5328-5341, 1985
A;Title: Amino acid sequence of the heavy chain of human alpha-factor XIIa (activated Ha
A;Reference number: A22248; MUID:85182674; PMID:3886654
```

```
A;Accession: A22248
A;Molecule type: protein
A;Residues: 20-379 <MCM>
R;Fujikawa, K.; McMullen, B.A.
J. Biol. Chem. 258, 10924-10933, 1983
A;Title: Amino acid sequence of human beta-factor XIIa.
A;Reference number: A21037; MUID:83291041; PMID:6604055
A;Accession: A21037
A;Molecule type: protein
A;Residues: 354-362,373-615 <FUJ>
R;Harris, R.J.; Ling, V.T.; Spellman, M.W.
J. Biol. Chem. 267, 5102-5107, 1992
A;Title: O-linked fucose is present in the first epidermal growth factor domain of facto
A;Reference number: A44606; MUID:92184750; PMID:1544894
A;Contents: annotation; carbohydrate binding site
C;Genetics:
C;Gene: GDB:F12
A;Cross-references: GDB:119892; OMIM:234000
A;Map position: 5q34-5qter
A;Introns: 19/3; 39/1; 72/2; 96/1; 133/1; 177/1; 212/1; 267/2; 340/1; 417/2; 463/1; 511/1
C;Complex: factor XII, prekallikrein, and HMW kininogen form a complex bound to anionic s
C;Function:
A;Description: factor XIIa catalyzes the proteolytic activation of plasminogen, plasma pr
ikrein
A;Pathway: blood coagulation; fibrinolysis
C;Superfamily: coagulation factor XII; EGF homology; fibronectin type I repeat homology;
C;Keywords: blood coagulation; fibrinolysis; glycoprotein; hydrolase; kringle; plasma; se
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-372,373-615/Product: coagulation factor XIIa, alpha form #status experimental <Al2>
F;47-88/Domain: fibronectin type II repeat homology <FB2>
F;98-130/Domain: EGF homology <EGl>
F;135-170/Domain: fibronectin type I repeat homology <1F1>
F;178-209/Domain: EGF homology <EG2>
F;217-295/Domain: kringle homology <KRG>
F;298-356/Region: proline-rich
F;354-362,373-615/Product: coagulation factor XIIa, beta form #status experimental <B12>
F;373-609/Domain: trypsin homology <TRY>
F;98-110,104-119,121-130,135-163,161-170,178-189,183-198,200-209,217-295,238-277,266-290,
F;109/Binding site: carbohydrate (Thr) (covalent) #status experimental
F;249,433/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;299,305,328,329,337/Binding site: carbohydrate (Thr) (covalent) #status predicted
F;308/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;412,461,563/Active site: His, Asp, Ser #status predicted

Query Match         8.0%;     Score 75.5;  DB 1;  Length 615;
Best Local Similarity  20.7%;    Pred. No. 30;
Matches   29;  Conservative  13;  Mismatches  35;  Indels  63;  Gaps   6;

Qy     1 CP-LQGSHALCTCCFQPM----------------------------PDRRAEREQDPRVAP 32

Db   121 CFPQHLTGNHCQXEKCFEPQLLRFFHKNEIWYRTEQAAVARCQCKGPDAHCQ-----RLAS 175

Qy    33 QQCAV----------CLQPFCHLYWGCTRTGYCGYGCLAPFCELNLGDKCLDGVLNNNSYESDI 84

Db   176 QACRTNPCLHGGRCLEVEGHRLCHCP---VGYTGPFCDVDTKASCYDG---------------- 220

Qy    85 LKNYLATRGLTWKNMLTESL 104

Db   221 ------RGLSYRGLARTTL 233

RESULT 15
T16840
hypothetical protein T10E10.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T16840
R;Geisel, C.
submitted to the EMBL Data Library, October 1995
A;Description: The sequence of C. elegans cosmid T10E10.
A;Reference number: Z18588
A;Accession: T16840
A;Status: preliminary; translated from GB/EMBL/DDBJ
```

```
Db   258 VFSYQRRMHSFSDCAWSTSC 277
```

RESULT 10
JC6565
four-and-a-half LIM-only protein 2 - human
C;Species: Homo sapiens (man)
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C;Accession: JC6565
R;Chan, K.K.; Tsui, S.K.W.; Lee, S.M.Y.; Luk, S.C.W.; Liew, C.C.; Fung, K.P.; Waye, M.M.
Gene 210, 345-350, 1998
A;Title: Molecular cloning and characterization of FHL2, a novel LIM domain protein pref
A;Reference number: JC6565
A;Accession: JC6565
A;Molecule type: mRNA
A;Residues: 1-279 <CHA>
A;Cross-references: GB:U29332; NID:g1845201; PIDN:AAC52073.1; PID:g1377897
A;Experimental source: heart
C;Genetics:
A;Map position: 2q12-2q13
C;Superfamily: LIM metal-binding repeat homology
C;Keywords: cardiac muscle; heart; zinc finger
F;221-275/Domain: LIM metal-binding repeat homology <LIM>

Query Match   8.4%;   Score 79;   DB 2;   Length 279;
Best Local Similarity 23.8%;   Pred. No. 6.5;
Matches 29;   Conservative 10;   Mismatches 51;   Indels 32;   Gaps 5;

Qy   9 LCTCFCQPMPDRRAEREQDPRVAPQQCAVCLQPPCHLYWGCTRTGCYGCLAPFCELNLGD 68
           ||| | : : :|| : | |||||
Db   187 VCTACRKQLSGQRFTARDD------FAYLNCFCDLY---AKKCAGCTNPISGLG-GT 234

Qy   69 KCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYC 128
            |||| |
Db   235 KYIS--FEERQWHNDCFNCKKCCSLSLVGRGFLTER-----------DDILCPD 274

Qy   129 CG 130
            ||
Db   275 CG 276
```

RESULT 11
S55194
DNA-directed DNA polymerase (EC 2.7.7.7) III regulatory chain - yeast (Saccharomyces cer
N;Alternate names: DNA-directed DNA polymerase delta small chain; HUS2 protein; HYS2 pro
C;Species: Saccharomyces cerevisiae
C;Date: 30-Oct-1998 #sequence_revision 30-Oct-1998 #text_change 21-Jul-2000
C;Accession: S55194; S57021; S59122
R;de Haan, M.; Smits, P.H.M.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A;Reference number: S55183
A;Accession: S55194
A;Molecule type: DNA
A;Residues: 1-487 <DEH>
A;Cross-references: EMBL:X87611; NID:g854567; PIDN:CAA60928.1; PID:g854579
R;de Haan, M.; Grivell, L.A.; Smits, P.H.M.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S56771
A;Accession: S57021
A;Molecule type: DNA
A;Residues: 1-487 <ZAG>
A;Cross-references: EMBL:Z49506; NID:g1015629; PIDN:CAA89528.1; PID:g1015630; GSPDB:GN00
R;Sugimoto, K.; Sakamoto, Y.; Takahashi, O.; Matsumoto, K.
Nucleic Acids Res. 23, 3493-3500, 1995
A;Title: HYS2, an essential gene required for DNA replication in Saccharomyces cerevisia
A;Reference number: S59122; MUID:96032843; PMID:7567461
A;Accession: S59122
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-155, 'H', 157-464, 'N', 466-487 <SUG>
A;Cross-references: EMBL:D50324; NID:g987711; PIDN:BAA08859.1; PID:g1256943
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995

C;Genetics:
A;Gene: SGD;HUS2, HYS2; MIPS;YIR006w
A;Cross-references: SGD:S0003766; MIPS:YIR006w
A;Map position: 10R
C;Complex: heterodimer of catalytic (see PIR:RNBYL3) and regulatory chain
C;Function:
A;Description: DNA polymerase delta is involved in DNA replication and in nucleotide exc
C;Superfamily: human DNA-directed DNA polymerase delta regulatory chain
C;Keywords: DNA repair; DNA replication; nucleotidyltransferase; nucleus

Query Match   8.1%;   Score 76.5;   DB 1;   Length 487;
Best Local Similarity 27.3%;   Pred. No. 19;
Matches 30;   Conservative 19;   Mismatches 52;   Indels 9;   Gaps 4;

```
Qy    54 CYGCLAPFCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNML-TESLVALQRGVF 112
           : : |||  |||| :        :       :|:
Db   104 CVGSI--YCEMKYKPNVLDEVINDTYGAPDLTKSYTDKEGGSDEIMLEDESGRVLLVGDF 161

Qy   113 LLSDYRVTGDTVLCYCGLR----SFRELTYQYRQNIPASELPVAVTSRP 158
            : |: |  |  :|:           : | : |    :|:
Db   162 IRSTPFITG--VVVGILGMEAEAGTFQVLDICYPTPLPQNFPAPIATCP 209
```

RESULT 12
A37926
fumarylacetoacetase (EC 3.7.1.2) - human
C;Species: Homo sapiens (man)
C;Date: 28-Feb-1992 #sequence_revision 28-Feb-1992 #text_change 21-Jul-2000
C;Accession: A37926; S09213
R;Phaneuf, D.; Labelle, Y.; Berube, D.; Arden, K.; Cavenee, W.; Gagne, R.; Tanguay, R.M.
Am. J. Hum. Genet. 48, 525-535, 1991
A;Title: Cloning and expression of the cDNA encoding human fumarylacetoacetate hydrolase
A;Reference number: A37926; MUID:91150763; PMID:1998338
A;Accession: A37926
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-419 <PHA>
R;Agsteribbe, E.; van Faassen, H.; Hartog, M.V.; Reversma, T.; Taanman, J.W.; Pannekoek,
Nucleic Acids Res. 18, 1887, 1990
A;Title: Nucleotide sequence of cDNA encoding human fumarylacetoacetase.
A;Reference number: S09213; MUID:90245581; PMID:2335361
A;Accession: S09213
A;Molecule type: mRNA
A;Residues: 71-419 <AGS>
A;Cross-references: EMBL:X51728; NID:g31290; PIDN:CAA36016.1; PID:g31291
C;Genetics:
A;Gene: GDB;FAH
A;Cross-references: GDB:119901; OMIM:276700
A;Map position: 15q23-15q25
C;Keywords: dimer; hydrolase

Query Match   8.0%;   Score 75.5;   DB 2;   Length 419;
Best Local Similarity 24.3%;   Pred. No. 21;
Matches 35;   Conservative 10;   Mismatches 54;   Indels 45;   Gaps 6;

```
Qy    17 MPDRRAEREQDPRVAPQQC----------------------------AVCLQPFCHLYW---G 48
           ||||| ||      :  |                            :  : |:
Db   270 MPFAVPNPKQDPRPLPYLCHDEPYTFDINLSVNLKGEGMSQAATICKSNFKYMYWTMLQQ 329

Qy    49 CTRTGCYGCLAPFCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQ 108
           |   : ||              |    : |:
Db   330 LTHHSVNG-----CNLRPGDLLASGTISGPEPE----NFGSMLELSWKGTKPIDLGNGQ 379

Qy   109 RGVFLLSDYRVTGDTVLC--YCCG 130
            |       |   |       ||||
Db   380 TRKFLLD-----GDEVIITGCYCG 398
```

RESULT 13
H84724
probable ARI-like RING zinc finger protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)

```
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: A90560
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-192 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33672.1; PID:g13359706; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0249
C;Superfamily: phosphoheptose isomerase

Query Match          8.5%; Score 80; DB 2; Length 192;
Best Local Similarity 30.6%; Pred. No. 3.6;
Matches   22; Conservative 15; Mismatches 27; Indels 8; Gaps 3;

Qy   83 DILKNYLATRGLTWKNMLTB--SLVALQRGVFLLSD-YRVTGDTVLC-----YCCGLRSF 134
        |:::||  |  ::|  ::||   |:||  ||| |  : ::|:|   :: |    ::|
Db    4 DLIRNELNEAAETLANFLKDDANIHAIQRAAVLLADSFKAGGKVLSCGNGGSHCDAMHFA 63

Qy  135 RELTYQYRQNIP 146
        ||| :||:|
Db   64 EELTGRYRENRP 75


RESULT 7
G85510
phosphoheptose isomerase [imported] - Escherichia coli (strain O157:H7, substrain EDL933
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: G85510
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: G85510
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-192 <STO>
A;Cross-references: GB:AB005174; NID:g12512990; PIDN:AAG54547.1; GSPDB:GN00145; UWGP:Z02
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: gmhA
C;Superfamily: phosphoheptose isomerase

Query Match          8.5%; Score 80; DB 2; Length 192;
Best Local Similarity 30.6%; Pred. No. 3.6;
Matches   22; Conservative 15; Mismatches 27; Indels 8; Gaps 3;

Qy   83 DILKNYLATRGLTWKNMLTB--SLVALQRGVFLLSD-YRVTGDTVLC-----YCCGLRSF 134
        |:::||  |  ::|  ::||   |:||  ||| |  : ::|:|   :: |    ::|
Db    4 DLIRNELNEAAETLANFLKDDANIHAIQRAAVLLADSFKAGGKVLSCGNGGSHCDAMHFA 63

Qy  135 RELTYQYRQNIP 146
        ||| :||:|
Db   64 EELTGRYRENRP 75


RESULT 8
AD0542
phosphoheptose isomerase [imported] - Salmonella enterica subsp. enterica serovar Typhi
C;Species: Salmonella enterica subsp. enterica serovar Typhi
C;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AD0542
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov


A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AD0542
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-192 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD08780.1; PID:g16501596; GSPDB:GN00176
C;Genetics:
A;Gene: STY0355
C;Superfamily: phosphoheptose isomerase

Query Match          8.5%; Score 80; DB 2; Length 192;
Best Local Similarity 30.6%; Pred. No. 3.6;
Matches   22; Conservative 15; Mismatches 27; Indels 8; Gaps 3;

Qy   83 DILKNYLATRGLTWKNMLTE--SLVALQRGVFLLSD-YRVTGDTVLC-----YCCGLRSF 134
 .      |:::||  |  ::|  ::||   |:||  ||| |  : ::|:|   :: |    ::|
Db    4 DLIRNELNEAAETLANFLKDDANIHAIQRAAVLLADSFKAGGKVLSCGNGGSHCDAMHFA 63

Qy  135 RELTYQYRQNIP 146
        ||| :||:|
Db   64 EELTGRYRENRP 75


RESULT 9
T00480
probable RING zinc finger protein At2g34990 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein F19I3.22
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 23-Mar-2001
C;Accession: T00480; C84763
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul,
submitted to the EMBL Data Library, April 1998
A;Description: Arabidopsis thaliana chromosome II BAC F19I3 genomic sequence.
A;Reference number: Z14160
A;Accession: T00480
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-302 <ROU>
A;Cross-references: EMBL:AC004238; NID:g3033373; PID:g3033394
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; N
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-302 <STO>
A;Cross-references: GB:AB002093; NID:g3033394; PIDN:AAC12838.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g34990; F19I3.22
A;Map position: 2
C;Superfamily: Arabidopsis hypothetical protein F19I3.22; RING finger homology

Query Match          8.5%; Score 80; DB 2; Length 302;
Best Local Similarity 22.5%; Pred. No. 5.7;
Matches   45; Conservative 18; Mismatches 53; Indels 84; Gaps 10;

Qy   34 QCAVCLQPF--------CHLY------WGCTRTGCYGCLAPFCELNLGDKCL---- 71
        |  |  ::             ::              |  |    |:  |  |
Db   95 ECAVCICEFEDHETLRLMPECHVFHADCVSVWLSDHSTC-----PLCRVDI---CLQPG 146

Qy   72 --------DGVLNNSYEBDILKNYLATRGLTWKNMLTES-----LVALQRGVFLLSD 116
                        |::|                |:|:|
Db  147 ERSYLNPEPDLVESTNSHLFD--------GVTWTNRNRPSRSWSTRLSQCRVSQILISR 197

Qy  117 YRVTGDTVLCYCCGLRSF--------RELTYQ--------YRQNIPASELP 151
         ||  |::|:|:|:|               :|:         || | |  |
Db  198 SHSTGHSVVQPLDNLDRFTLRLPEEVRRQLTKKTVDNVAFSQARSSRRGYRSRSAGSERS 257

Qy  152 V-----AVTSRPDCYWGRNC 166
```

A;Accession: C96516
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-473 <STO>
A;Cross-references: GB:AE005173; NID:g5668816; PIDN:AAD46042.1; GSPDB:GN00141
C;Genetics:
A;Gene: F16N3.15
A;Map position: 1

```
Query Match         14.6%;   Score 137.5; DB 2;  Length 473;
Best Local Similarity   25.0%;   Pred. No. 4.2e-05;
Matches   45;  Conservative   26;  Mismatches   88;  Indels   21;  Gaps    6;

Qy       1 CPLQGSHALCTCCFQPMDPRRAAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
             ||:|| |: |     ||||||||||||||||||:||||||||: ::      : :
Db     275 CEHHGAHLQCHLCQGMMPFRANLQ------VPLHCKGCDRPFCGAYWS-SENVTQGVSGP 327

Qy      61 FCELN---LGDKCLDGV---LNNNSYESDILKNYLATRGLTWKNMLTESL-----VAL 107
             ||::|   :|:: : |    ::  :| :::  ||: |:|::::::|
Db     328 VCVRETFRPISERTITRIPPITHEMNRHEQDITQRCIAHMEKTVPDVVAEWLRLFNNREI 387

Qy     108 QRGVFLLSDYR-VTGDTVLCYCCGLRSFRELTYQRQNIPASELPVAVTSRPDCYWGRNC 166
             |||||| |:|  ||||||||:| |:||||:|||:::| |::::  ::| : ::: |
Db     388 DRSRMPLNHAEBTITASTHVCNDCYDKLVGFLLIWFRITLPRNHLPADVAAREDCWYGVAC 447
```

RESULT 3
T22759
hypothetical protein F55H12.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C;Accession: T22759
R;Dobson, R.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19610
A;Accession: T22759
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2824 <WIL>
A;Cross-references: EMBL:Z81091; PIDN:CAB03143.1; GSPDB:GN00019; CESP:F55H12.3
A;Experimental source: clone F55H12
C;Genetics:
A;Gene: CESP:F55H12.3
A;Map position: 1
A;Introns: 87/1; 98/1; 126/2; 201/3; 343/3; 406/1; 576/3; 656/1; 825/3; 869/1; 909/1; 96
7/1; 1755/2; 1800/1; 1850/3; 1896/1; 2003/3; 2035/3; 2082/3; 2119/1; 2144/1; 2200/2; 227
C;Superfamily: LDL receptor ligand-binding repeat homology
F;243-279/Domain: LDL receptor ligand-binding repeat homology <LDL>

```
Query Match          8.9%;   Score 83.5; DB 2;  Length 2824;
Best Local Similarity   27.1%;   Pred. No. 26;
Matches   55;  Conservative   12;  Mismatches   51;  Indels   85;  Gaps   15;

Qy      10 CTCCFQP------MPDRRAAEREQDPRVAPQQCAVCLQ------------------P 41
             ||||||                       ||||||||||
Db    1407 CLCNPQPPCPSTELLPQR-DQIGDCTLNSCQNGVCLQMNGMHDCACNEYFIVDRENGTP 1464

Qy      42 FCH---LYWGCTRTGCYGC------LAPFCELN-LGDKC-LDGVLNNNSYE---SD 83
             ||    ||| |:|  |  |       ||   :|  :||:   | ::::|:||
Db    1465 FCKPNHCLFTNQTRNSGYDCSTGRENALFLCP--NLNEFGDFCQYEGQLYNNSYTYILSA 1522

Qy      84 ILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQRQ 143
             |||||||||:        ::|:||:||||| :|||  |||||||:| |:||||:|||:|
Db    1523 GIENAIATN-------LCESLDTL--------YAVPN----TFCV----------Q 1549

Qy     144 NIPASELPVAVTSRPDCYWGRNC 166
             ::  |
Db    1550 N-PTSTTP-SIHRCDFCYGGENC 1570
```

RESULT 4
T31070
notch homolog - sea urchin (Lytechinus variegatus)

C;Species: Lytechinus variegatus (variegated urchin)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
C;Accession: T31070
R;Sherwood, D.R.; McClay, D.R.
Development 124, 3363-3374, 1997
A;Title: Identification and localization of a sea urchin Notch homologue: insights into
A;Reference number: Z20966; MUID:97454256; PMID:9310331
A;Accession: T31070
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2531 <SHE>
A;Cross-references: EMBL:AF000634; NID:g2570350; PID:g2570351; PIDN:AAB82088.1
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology

```
Query Match          8.8%;   Score 83; DB 2;  Length 2531;
Best Local Similarity   28.6%;   Pred. No. 25;
Matches   30;  Conservative    9;  Mismatches   40;  Indels   26;  Gaps    6;

Qy       5 GSHALCTCCFQP--MPDRRAAEREQDPRVAPQQCAVCLQPFCHLYWGCTR-TGCYGCL--- 58
             | :|  |||    :||||||||||||||||||||||:|| ||: | ::    ||| |
Db     840 GYSCRCTSGFQGNFCDDDRNE--------------------CLFSPCRNGGSCTNLEGSFECSCLP 885

Qy      59 --APFCELNLGDKCLDGVLNNSYESDILKNYLAT--RGLTWKN 98
             | ||  |  |: ||  :::::||:: |||||:||
Db     886 GYDGPICEINI-DECASGPCTNGGICTDLIDDYFCSCQRGFTGKN 929
```

RESULT 5
G64746
phosphoheptose isomerase (EC 5.-.-.-) gmhA - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002
C;Accession: G64746
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G64746
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-192 <BLAT>
A;Cross-references: GB:AE000131; GB:U00096; NID:g1786415; PIDN:AAC73326.1; PID:g1786416;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: gmhA
C;Function:
A;Description: synthesis of glyceromannoheptose 7-phosphate
A;Pathway: inner core lipopolysaccharide biosynthesis
C;Superfamily: phosphoheptose isomerase
C;Keywords: isomerase

```
Query Match          8.5%;   Score 80; DB 1;  Length 192;
Best Local Similarity   30.6%;   Pred. No. 3.6;
Matches   22;  Conservative   15;  Mismatches   27;  Indels    8;  Gaps    3;

Qy      83 DILKNYLATRGLTWKNMLTE--SLVALQRGVFLLSD-YRVTGDTVLC----YCCGLRSF 134
             :|||:|||:||||||||:|   :||||||||:||:|
Db       4 DLIRNELNEAAETIANFLKDDANIHAIQRAAVLLADSFKAGGKVLSCGNGGSHCDAMHFA 63

Qy     135 RELTYQYRQNIP 146
             ::||:|:|:|
Db      64 EELTGRYRENRP 75
```

RESULT 6
A90660
phosphoheptose isomerase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: A90660
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

```
                    GenCore version 5.1.6
         Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:       May 7, 2004, 14:41:17 ; Search time 9.29039 Seconds
                              (without alignments)
                              1718.743 Million cell updates/sec

Title:        US-10-048-046-2_COPY_476_641
Perfect score: 943
Sequence:     1 CPLQGSHALCTCCFQPMPDR.........ASELPVAVTSRPDCYWGRRNC 166

Scoring table: BLOSUM62
              Gapop 10.0 , Gapext 0.5

Searched :    283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:        283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                Maximum Match 100%
                Listing first 45 summaries

Database :    PIR_78:*
              1: _pir1:*
              2: _pir2:*
              3: _pir3:*
              4: _pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                       SUMMARIES
```

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	943	100.0	306	2	T46399	hypothetical prote
2	137.5	14.6	473	2	C96516	F16N3.15 [imported
3	83.5	8.9	2824	2	T22759	hypothetical prote
4	83	8.8	2531	2	T31070	notch homolog - se
5	80	8.5	192	1	G64746	phosphoheptose iso
6	80	8.5	192	2	A90660	phosphoheptose iso
7	80	8.5	192	2	G85510	phosphoheptose iso
8	80	8.5	192	2	AD0542	phosphoheptose iso
9	80	8.5	302	2	T00480	probable RING zinc
10	79	8.4	279	2	JC6565	four-and-a-half LI
11	76.5	8.1	487	1	S55194	DNA-directed DNA p
12	75.5	8.0	419	2	A37926	fumarylacetoacetas
13	75.5	8.0	543	2	H84724	probable ARI-like
14	75.5	8.0	615	1	KFHU12	coagulation factor
15	75.5	8.0	1101	2	T16840	hypothetical prote
16	75	8.0	193	2	AB0394	phosphoheptose iso
17	75	8.0	1458	2	A45665	adult-specific bru
18	74.5	7.9	419	2	A56825	fumarylacetoacetas
19	74.5	7.9	603	2	S28941	coagulation factor
20	73.5	7.8	362	2	T02605	hypothetical prote
21	73.5	7.8	406	2	T24021	hypothetical prote
22	73.5	7.8	419	2	A42219	fumarylacetoacetat
23	73.5	7.8	419	2	JH0467	fumarylacetoacetas
24	73.5	7.8	419	2	A42219	fumarylacetoacetat
25	73	7.7	679	2	B96599	protein F20N2.12 [
26	73	7.7	603	2	A36630	prostaglandin-endo
27	72	7.6	3075	2	S14458	laminin alpha-1 ch
28	71.5	7.6	483	2	S27880	Nasopressin recept
29	71.5	7.6	4550	2	T18440	hypothetical prote
—	71	7.5	3635	2	T10053	laminin alpha 5 ch

30	70.5	7.5	192	2	E70414	hypothetical prote
31	70.5	7.5	351	2	D64409	iron(III) dicitrat
32	70.5	7.5	387	2	JC5556	adhalin - mouse
33	70.5	7.5	421	2	AC1096	hypothetical prote
34	70.5	7.5	1135	1	JQ1928	G2-G1 polyprotein
35	70	7.4	661	2	B96696	hypothetical prote
36	70	7.4	802	2	T24293	hypothetical prote
37	70	7.4	949	2	T24294	hypothetical prote
38	69.5	7.4	320	2	T45909	ribulose-bisphosph
39	69.5	7.4	401	2	T02958	crumbs protein - f
40	69.5	7.4	2139	2	A35672	probable ROK-famil
41	69	7.3	302	2	AE0645	hypothetical prote
42	69	7.3	347	2	T33204	probable RING zinc
43	69	7.3	378	2	T00481	hypothetical prote
44	69	7.3	456	2	T19817	hypothetical prote
45	68.5	7.3	444	2	A55071	hydrogen peroxide-

```
                       ALIGNMENTS

RESULT 1
T46399
hypothetical protein DKFZp434N2420.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46399
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23031
A;Accession: T46399
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-306 <AAA>
A;Cross-references: EMBL:AL137561
A;Experimental source: adult testis; clone DKFZp434N2420
C;Genetics:
A;Note: DKFZp434N2420.1

Query Match             100.0%;   Score 943;   DB 2;   Length 306;
Best Local Similarity   100.0%;   Pred. No. 7e-80;
Matches  166;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      118 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 177

Qy       61 FCELNLGDKLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      178 FCELNLGDKLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 237

Qy      121 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
            ||||||||||||||||||||||||||||||||||||||||||||||
Db      238 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 283

RESULT 2
C96516
F16N3.15 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: C96516
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
```

```
;   APPLICANT:  Watanabe,Colin K
;   APPLICANT:  Wood,William
;   APPLICANT:  Zhang, Zemin
;   TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;   TITLE OF INVENTION: ACIDS ENCODING THE SAME
;   FILE REFERENCE: P3330R1C30
;   CURRENT APPLICATION NUMBER: US/10/123,155
;   CURRENT FILING DATE:  2002-04-15
;   Prior Application removed - See Palm or File Wrapper
;   NUMBER OF SEQ ID NOS: 550
;   SEQ ID NO 495
;     LENGTH: 771
;     TYPE: DNA
;   ORGANISM: Homo Sapien
US-10-123-155-495

Query Match          8.2%;  Score 77;  DB 14;  Length 771;
Best Local Similarity   29.4%;  Pred. No. 32;
Matches  20;  Conservative  1;  Mismatches  29;  Indels  18;  Gaps   2;

Qy      1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTG-CYGCLA 59
            | |:|   ||||||          |    |  | ||:|     ||| | ||
Db     98 CCCAGTGGCCCCC-----------------ATGACTCCTTACCTGATGCTGTGCCAGCCA 140

Qy     60 PFCELNLG 67
            ||||:|
Db    141 CACAAGAG 148

Search completed: May 7, 2004, 15:06:52
Job time : 27.2936 secs
```

```
Qy   89 L--ATRGLTW-KMLTESLVALQRGVFLLSDYRVTGDTVL-----CYCCGLRSFR 135
           :|    ||       |:        ::    |:       ||::|  |:
Db  244 LGHSRASVMWNRTQVVGILVGL--GIIAL----VTSPILLLASPCIICCVCKSCR 292
```

RESULT 14
US-10-142-426-495
; Sequence 495, Application US/10142426
; Publication No. US20040048333A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C224
; CURRENT APPLICATION NUMBER: US/10/142,426
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
SEQ ID NO 495
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-142-426-495

Query Match   8.2%; Score 77; DB 12; Length 771;
Best Local Similarity 29.4%; Pred. No. 32;
Matches   20; Conservative   1; Mismatches   29; Indels   18; Gaps   2;

```
Qy    1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTG-CYGCLA 59
          |||:||              |  |||||||||||||          |||||||
Db   98 CCCAGTGGCCCCC------------------ATGACTCCTTACCTGATGCTGTGCCAGCCA 140

Qy   60 PFCELNLG 67
          ||||||
Db  141 CACAAGAG 148
```

RESULT 15
US-10-123-155-495
; Sequence 495, Application US/10123155
; Publication No. US20030066794A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel

RESULT 12
US-10-425-114-62319
; Sequence 62319, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou,Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 62319
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4666-024-HS_FLI.pep
US-10-425-114-62319

Query Match   8.3%; Score 78.5; DB 12; Length 216;
Best Local Similarity 24.8%; Pred. No. 4.4;
Matches   40; Conservative   11; Mismatches   55; Indels   55; Gaps   10;

```
Qy    2 PLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPF 61
          |||||:||      |:|                              :|
Db   74 PTQWSSGICAC----FDD----------PQSC----------------CIGAICP-   98

Qy   62 CELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTG 121
                   :  |:         |:|||| :|:  ||       :|::
Db   99 --------CFLFGKNAQFLGSGTLAGSCTTHCMLW-GLLT-SLCCVFTGGLVLA---VPG 145

Qy  122 DTVLCYCCGLRSFRELTYQYRQNIPASELPVA-VTSRPDCY 161
           :|::|  |:      |:|            ||  :|:
Db  146 SAVACVACGYRSALRTKY---NLP--EAPCGDLTTHLFCH 180
```

RESULT 13
US-10-108-260A-3965
; Sequence 3965, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3965
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3965

Query Match   8.2%; Score 77; DB 15; Length 303;
Best Local Similarity 22.3%; Pred. No. 9.9;
Matches   39; Conservative   19; Mismatches   65; Indels   52; Gaps   10;

```
Qy    1 CPLQGSH-----------ALCTCCFQPMPDRRAEREQDPRVAP------------ 32
          ||:|                ||  |:                  :|
Db  130 CPVASSDPGQPVLVBCPSCHLKFCSCCKDAWHAEVSCRDSQPIVLPTEHRALFGTDAEAP 189

Qy   33 -QQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELNLGDKCLDGVLNN--NSYESDILKNY 88
          ||| ||:   ::              :|:          |:           |:
Db  190 IKQCPVC-RYYIERNEGCAQMMCKNCKHTFCWY-----CLQNLDNDIFLRHYDKGPCRNK 243
```

US-09-942-052-731

Query Match 9.0%; Score 85; DB 10; Length 229;
Best Local Similarity 26.6%; Pred. No. 0.96;
Matches 38; Conservative 15; Mismatches 52; Indels 38; Gaps 6;

Qy   26 QDPRVAPQQCAVC--LQPFCHLYWGCTRTGCYGCLAPFCELNLGDKCLDGVLNNNSYES 82
             :|| ||| :||
Db   71 QPERCAVFQCAQCHAVLADSVHLAWDLSR-----------SLGAVVFSRVTNNVVLEA 117

Qy   83 DILKNYLAT-RGLTWKNML---------------TESLVALQRGVFLLSDYRVTGDTVLC 126
             : : : :|:   :|
Db   118 PFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHAALLAALRGHFCLS------SDKMVC 171

Qy   127 YCCGLRSFRELTYQYRQNIPASE 149
             :| :||
Db   172 YLLKTKAIVNASEMDIQNVPLSE 194

RESULT 9
US-10-424-599-184701
; Sequence 184701, Application US/10424599
; Publication No. US2004003107ZA1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184701
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(202)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_137800C.1.pep
US-10-424-599-184701

Query Match 8.9%; Score 83.5; DB 12; Length 202;
Best Local Similarity 20.1%; Pred. No. 1.2;
Matches 37; Conservative 19; Mismatches 65; Indels 63; Gaps 7;

Qy   4 QGSHALCTCFCQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCE 63
             |||||| ||
Db   42 ------------MGQSCMLLPYPVTSFXY---KGCIGCLCP--- 78

Qy   64 LNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDT 123
             ||||| : :   :|
Db   79 ------CFLFGKNADFLGSGTFLGSCVTHFILW-SVVNTACCLLTDGLF----WGLPGCL 127

Qy   124 VLCYCCGLRSFRELTY-----------------QYRQ------NIPASELPVAV 154
             :||| :|:   :|:
Db   128 VSCYACGYRKALRSKYNLPEAPCGDFVTHFCCHPCAICQEYREIRERSGDCEATDLKLAV 187

Qy   155 TSRP 158
             :|:
Db   188 VTAP 191

RESULT 10
US-10-369-493-5166
; Sequence 5166, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5166
; LENGTH: 2824
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5166

Query Match 8.9%; Score 83.5; DB 15; Length 2824;
Best Local Similarity 27.1%; Pred. No. 34;
Matches 55; Conservative 12; Mismatches 51; Indels 85; Gaps 15;

Qy   10 CTCCFQP------MPDRRAEREQDPRVAPQQCAVCLQ----------P 41
             :|| :|
Db 1407 CLCNPQPPCPSTELLPQR--DQIGDCTLNSCQNGVCLQMNGMHDCACNEYFIVDRENGTP 1464

Qy   42 FCH---LYWGCTRTGCYGC---------LAPFCELN-LGDKC-LDGVLNNNSYE---SD 83
             || ||| | :| | ||||||||| :|
Db 1465 FCKPNHCLFTNQTRNSGYDCSTGRENALFLCP--NLNEFGDFCQYEGGLYNNSYTYILSA 1522

Qy   84 ILKNYLAIRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQ 143
             : |
Db 1523 GIENAIAIN------LCESLDTL--------YAVPN----TFCV-----------Q 1549

Qy   144 NIPASELPVAVTSRPDCYWGRNC 166
             | :   : || | :||
Db 1550 N-PTSTTP-SIHRCDFCYGGENC 1570

RESULT 11
US-10-424-599-224914
; Sequence 224914, Application US/10424599
; Publication No. US2004003107ZA1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 224914
; LENGTH: 129
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_45128C.1.pep
US-10-424-599-224914

Query Match 8.7%; Score 82; DB 12; Length 129;
Best Local Similarity 19.1%; Pred. No. 0.97;
Matches 22; Conservative 26; Mismatches 31; Indels 36; Gaps 4;

Qy   15 QPMPDRRAEREQDPRVAPQQCAVCLQPF----------------CHLYWGCTRTGCYG 56
             :|: :: :: :|: :| : : |:
Db   22 KPLKEKQRENDEDSK----SCAICLEDFDPNEEVMLTPCNHMFHEDCIVPWLTSKGQCPV 77

Qy   57 CLAPFCELNLGDKCLDGVLNNNSYESDILKNYLiATRGLTWKNMLTESLVALQRGV 111
             | :   ::::   : :|:   : :   ::
Db   78 CRFVICEIGRG--------NHSSFNNNDIAN-----LEPSNLINGELLSILRAM 118

```
;   PRIOR APPLICATION NUMBER: 60/228,432
;   PRIOR FILING DATE: 2000-08-28
;   SOFTWARE: PatentIn Ver. 2.1
;   NUMBER OF SEQ ID NOS: 744
;   SEQ ID NO 728
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Unknown Organism
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism: 85P1B3/OIP5
;   OTHER INFORMATION: clone A protein
US-09-942-052-728

Query Match             9.0%;  Score 85;  DB 10;  Length 229;
Best Local Similarity  26.6%;  Pred. No. 0.96;
Matches  38;  Conservative  15;  Mismatches  52;  Indels  38;  Gaps  6;

Qy    26 QDPRVAPQQCAVC--LQPFCHLYWGCTRTGCYGCLAPFCELNLGDKCLDGVLNNSYES 82
            |     ||||  :|   |        |    |     |  |      :
Db    71 QPERCAVFQCAQCHAVLADSVHLAWDLSR----------SLGAVVFSRVTNNVVLEA 117

Qy    83 DILKNYLAT-RGLTWKNML-------------TESLVALQRGVFLLSDYRVTGDTVLC 126
            :: |: |:|                     |  : :|: ||
Db   118 PFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHAALAALRGHFCLS-----SDKMVC 171

Qy   127 YCCGLRSFRELTYQYRQNIPASE 149
            |       :       ||: ||
Db   172 YLLKTKAIVNASEMDIQNVPLSE 194


RESULT 7
US-09-942-052-730
;   Sequence 730, Application US/09942052
;   Publication No. US20030170626A1
;   GENERAL INFORMATION:
;   APPLICANT: Raitano, Arthur B.
;   APPLICANT: Faris, Mary
;   APPLICANT: Hubert, Rene S.
;   APPLICANT: Afar, Daniel
;   APPLICANT: Ge, Wangmao
;   APPLICANT: Challita-Eid, Pia M.
;   TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
;   TITLE OF INVENTION: USEFUL IN TREATMENT AND DETECTION OF CANCER
;   FILE REFERENCE: 51158-20028.00
;   CURRENT APPLICATION NUMBER: US/09/942,052
;   CURRENT FILING DATE: 2001-08-28
;   PRIOR APPLICATION NUMBER: 60/228,432
;   PRIOR FILING DATE: 2000-08-28
;   NUMBER OF SEQ ID NOS: 744
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 730
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Unknown Organism
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism: 85P1B3 protein
;   OTHER INFORMATION: sequence
US-09-942-052-730

Query Match             9.0%;  Score 85;  DB 10;  Length 229;
Best Local Similarity  26.6%;  Pred. No. 0.96;
Matches  38;  Conservative  15;  Mismatches  52;  Indels  38;  Gaps  6;

Qy    26 QDPRVAPQQCAVC--LQPFCHLYWGCTRTGCYGCLAPFCELNLGDKCLDGVLNNSYES 82
            |     ||||  :|   |        |    |     |  |      :
Db    71 QPERCAVFQCAQCHAVLADSVHLAWDLSR----------SLGAVVFSRVTNNVVLEA 117

Qy    83 DILKNYLAT-RGLTWKNML-------------TESLVALQRGVFLLSDYRVTGDTVLC 126
            :: |: |:|                     |  : :|: ||
Db   118 PFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHAALAALRGHFCLS-----SDKMVC 171

Qy   127 YCCGLRSFRELTYQYRQNIPASE 149
            |       :       ||: ||
Db   172 YLLKTKAIVNASEMDIQNVPLSE 194


RESULT 8
US-09-942-052-731
;   Sequence 731, Application US/09942052
;   Publication No. US20030170626A1
;   GENERAL INFORMATION:
;   APPLICANT: Raitano, Arthur B.
;   APPLICANT: Faris, Mary
;   APPLICANT: Hubert, Rene S.
;   APPLICANT: Afar, Daniel
;   APPLICANT: Ge, Wangmao
;   APPLICANT: Challita-Eid, Pia M.
;   TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
;   TITLE OF INVENTION: USEFUL IN TREATMENT AND DETECTION OF CANCER
;   FILE REFERENCE: 51158-20028.00
;   CURRENT APPLICATION NUMBER: US/09/942,052
;   CURRENT FILING DATE: 2001-08-28
;   PRIOR APPLICATION NUMBER: 60/228,432
;   PRIOR FILING DATE: 2000-08-28
;   NUMBER OF SEQ ID NOS: 744
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 731
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Homo sapiens
```

```
Query Match          39.9%;  Score 376; DB 15;  Length 349;
Best Local Similarity 98.4%;  Pred. No. 1.1e-31;
Matches   61;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  276 CPLQGSHALCTCCFQPMPDRRVREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 335

Qy   61 FC 62
          ||
Db  336 FC 337
```

RESULT 4
US-10-424-599-264601
; Sequence 264601, Application US/10424599
; Publication No. US2004031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 264601
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_80956C.1.pep
US-10-424-599-264601

```
Query Match          16.0%;  Score 150.5; DB 12;  Length 287;
Best Local Similarity 25.1%;  Pred. No. 1.2e-07;
Matches   46;  Conservative  26;  Mismatches  84;  Indels  27;  Gaps   6;

Qy    1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYW---GCTRTGCYGC 57
          ||  ||||   :          :   :  :  :   :          |:
Db   90 CKYDTVHLQCCACGGMPSRTGFG------IPQYCSGCDRSFCGAYWHALGVTGNGSY-- 141

Qy   58 LAPFCELN----LGDKCLDGV---LNNNSYESDILKNYLATRGLTWKNMLTESLVALQ- 108
            :   |:   |   :   |:     :    :  ::   :|:
Db  142 --PVCSQDTLRPISDHSIRIPLLAHEKNLHEQNITDSCIRQMGRTLPDVISEWIAKFEN 199

Qy  109 ----RGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWG 163
            :: :  ||||  ::  .   :   :     :        :  |
Db  200 REIDRRRMLNHAEMITARTFVCQDCYHKLVSFLLYWFRLSIPKHLLPPDESAREDCWYG 259

Qy  164 RNC 166
          |
Db  260 YAC 262
```

RESULT 5
US-09-942-052-728
; Sequence 728, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Afar, Daniel
; APPLICANT: Ge, Wangmao
; APPLICANT: Challita-Eid, Pia M.
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
; TITLE OF INVENTION: USEFUL IN TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 51158-20028.00
; CURRENT APPLICATION NUMBER: US/09/942,052
; CURRENT FILING DATE: 2001-08-28

```
Db  596 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 641
```

RESULT 2
US-10-094-749-1799
; Sequence 1799, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1799
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1799

```
Query Match          46.4%;  Score 438; DB 15;  Length 128;
Best Local Similarity 98.8%;  Pred. No. 6.7e-39;
Matches   80;  Conservative   1;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   86 KNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNI 145
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 QNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNI 84

Qy  146 PASELPVAVTSRPDCYWGRNC 166
          |||||||||||||||||||||
Db   85 PASELPVAVTSRPDCYWGRNC 105
```

RESULT 3
US-10-108-260A-4516
; Sequence 4516, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US2004005560A1e1 full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4516
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4516

QM protein - protein search, using sw model

```
          GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:    May 7, 2004, 14:49:13 ; Search time 26.2936 Seconds
                                    (without alignments)
                                    1752.371 Million cell updates/sec

Title:         US-10-048-046-2_COPY_476_641
Perfect score: 943
Sequence:    1 CPLQGSHALCTCCFQPMPDR.........ASELPVAVTSRPDCYWGRNC 166

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:  1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters:     1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : Published Applications AA:*
 1: /cgn2_6/ptcodata/2/pubpaa/US07_PUBCOMB.pep:*
 2: /cgn2_6/ptcodata/2/pubpaa/PCT_NEW_PUB.pep:*
 3: /cgn2_6/ptcodata/2/pubpaa/US06_NEW_PUB.pep:*
 4: /cgn2_6/ptcodata/2/pubpaa/US06_PUBCOMB.pep:*
 5: /cgn2_6/ptcodata/2/pubpaa/US07_NEW_PUB.pep:*
 6: /cgn2_6/ptcodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
 7: /cgn2_6/ptcodata/2/pubpaa/US08_NEW_PUB.pep:*
 8: /cgn2_6/ptcodata/2/pubpaa/US08_PUBCOMB.pep:*
 9: /cgn2_6/ptcodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptcodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptcodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptcodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptcodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptcodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptcodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptcodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptcodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptcodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	943	100.0	664	9	US-09-780-525-2	Sequence 2, Appli
2	438	46.4	128	15	US-10-094-749-1799	Sequence 1799, Ap
3	376	39.9	349	12	US-10-108-260A-4516	Sequence 4516, Ap
4	150.5	16.0	287	12	US-10-424-599-264601	Sequence 264601,
5	85	9.0	229	10	US-09-942-052-728	Sequence 728, App
6	85	9.0	229	10	US-09-942-052-729	Sequence 729, App
7	85	9.0	229	10	US-09-942-052-730	Sequence 730, App
8	85	9.0	229	10	US-09-942-052-731	Sequence 731, App
9	83.5	8.9	202	12	US-10-424-599-184701	Sequence 184701,
10	83.5	8.7	2824	15	US-10-369-493-5166	Sequence 5166, Ap
11	82	8.7	129	12	US-10-424-599-224914	Sequence 224914,
12	78.5	8.3	216	12	US-10-425-114-62319	Sequence 62319, A
13	77	8.2	303	15	US-10-108-260A-3965	Sequence 3965, Ap
14	77	8.2	771	12	US-10-142-426-495	Sequence 495, App
15	77	8.2	771	14	US-10-123-155-495	Sequence 495, App
16	77	8.2	771	14	US-10-146-731-495	Sequence 495, App
17	77	8.2	771	14	US-10-140-472-495	Sequence 495, App
18	77	8.2	771	14	US-10-141-761-495	Sequence 495, App
19	77	8.2	771	14	US-10-142-885-495	Sequence 495, App
20	77	8.2	771	14	US-10-158-790-495	Sequence 495, App
21	77	8.2	771	15	US-10-137-871-495	Sequence 495, App
22	77	8.2	771	15	US-10-140-923-495	Sequence 495, App
23	77	8.2	771	15	US-10-141-756-495	Sequence 495, App
24	77	8.2	771	15	US-10-141-759-495	Sequence 495, App
25	77	8.2	771	15	US-10-140-805-495	Sequence 495, App
26	77	8.2	771	15	US-10-140-864-495	Sequence 495, App
27	76.5	8.1	487	15	US-10-369-493-22178	Sequence 22178, A
28	76	8.1	567	12	US-10-369-022-2	Sequence 2, Appli
29	75.5	8.0	419	14	US-10-354-358-76	Sequence 76, Appl
30	75.5	8.0	615	10	US-09-858-909-2	Sequence 2, Appli
31	75.5	8.0	615	14	US-10-172-712-30	Sequence 30, Appl
32	75.5	8.0	615	16	US-10-449-132-2	Sequence 2, Appli
33	75	8.0	1835	10	US-09-457-571-15	Sequence 15, Appl
34	74.5	7.9	246	15	US-10-104-047-3660	Sequence 3660, Ap
35	74.5	7.9	2849	12	US-10-142-426-371	Sequence 371, App
36	74.5	7.9	2849	14	US-10-123-155-371	Sequence 371, App
37	74.5	7.9	2849	14	US-10-146-731-371	Sequence 371, App
38	74.5	7.9	2849	14	US-10-140-472-371	Sequence 371, App
39	74.5	7.9	2849	14	US-10-141-761-371	Sequence 371, App
40	74.5	7.9	2849	14	US-10-142-885-371	Sequence 371, App
41	74.5	7.9	2849	14	US-10-158-790-371	Sequence 371, App
42	74.5	7.9	2849	15	US-10-137-871-371	Sequence 371, App
43	74.5	7.9	2849	15	US-10-140-923-371	Sequence 371, App
44	74.5	7.9	2849	15	US-10-141-756-371	Sequence 371, App
45	74.5	7.9	2849	15	US-10-141-759-371	Sequence 371, App

ALIGNMENTS

```
RESULT 1
US-09-780-525-2
Sequence 2, Application US/09780525
; Patent No. US2002004223A1
; GENERAL INFORMATION:
;  APPLICANT: Bin-Bing Zhou
;  APPLICANT: Yuan Zhu
;  APPLICANT: Priya Chaturvedi
;  APPLICANT: Mark R. Hurle
;  APPLICANT: Xiaotong Li
;  TITLE OF INVENTION: FHAR1, A NEW RING FINGER PROTEIN
;  FILE REFERENCE: GP-7066B-C1
;  CURRENT APPLICATION NUMBER: US/09/780,525
;  CURRENT FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 09/456,876
;  PRIOR FILING DATE: 1999-12-08
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 664
;   TYPE: PRT
;   ORGANISM: HOMO SAPIENS
US-09-780-525-2

Query Match             100.0%;  Score 943;  DB 9;  Length 664;
Best Local Similarity   100.0%;  Pred. No. 4e-92;
Matches  166;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  476 CPLQGSHALCTCCFQPMPDRRABREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 535

Qy   61 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  536 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 595

Qy  121 GDTVLCYCCGLRSFRELTYQRQNIPASELPVAVTSRPDCYWGRNC 166
          |||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 15
US-09-798-051-9

; Sequence 9, Application US/09798051
; Patent No. 6632780
; GENERAL INFORMATION:
;   APPLICANT: Zhang, Ke
;   APPLICANT: Cam, Linh
;   APPLICANT: Nakayama, Naoki
;   TITLE OF INVENTION: Chordin-Like-2 Molecules and Uses Thereof
;   FILE REFERENCE: 01-005
;   CURRENT APPLICATION NUMBER: US/09/798,051
;   CURRENT FILING DATE: 2001-03-05
;   NUMBER OF SEQ ID NOS: 21
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 452
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-798-051-9

Query Match   7.2%; Score 67.5; DB 4; Length 452;
Best Local Similarity   25.0%; Pred. No. 27;
Matches   40; Conservative   18; Mismatches   73; Indels   29; Gaps   9;

Qy   4 QGSHALCTCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWG--CTRTGC--YGCLA 59
          | |:| || | || | |: |: |:| ||:
Db   23 QVKHSETYCMFQDKKYRVGER-WHPYLEPYGLVYCVNCICSENGNVLCSRVRCPNVHCLS 81

Qy   60 P-----FCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLL 114
          | : | | |:|: || || |:|
Db   82 PVHIPHLCCPRCPEDSLPPVNNKVTSKS-----CBYNGTTYQH--GELFVA--EGLF-- 129

Qy   115 SDYRVTGDTVLCYC------CGLRSFRELTYQYRQNIPAS 148
          | | |||: :||| : : ::|
Db   130 -QNRQPNQCTQCSCSSEGNVYCGLKTCPKLTCAFPVSVPDS 168

Search completed: May 7, 2004, 14:51:24
Job time : 12.0433 secs

```
; Sequence 5, Application US/09798051
; Patent No. 6632780
; GENERAL INFORMATION:
;   APPLICANT: Zhang, Ke
;   APPLICANT: Cam, Linh
;   APPLICANT: Nakayama, Naoki
;   TITLE OF INVENTION: Chordin-like-2 Molecules and Uses Thereof
;   FILE REFERENCE: 01-005
;   CURRENT APPLICATION NUMBER: US/09/798,051
;   CURRENT FILING DATE: 2001-03-05
;   NUMBER OF SEQ ID NOS: 21
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 429
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-798-051-5

Query Match          7.2%;    Score 68;  DB 4;  Length 429;
Best Local Similarity  24.3%;    Pred. No. 22;
Matches   25;  Conservative   8;  Mismatches  46;  Indels  24;  Gaps   3;

Qy    2 PLQGSHALCTCCFQP------MPDRRAE-----------------REQDPRVAPQQCAVCLQ 40
               |:|    : |.:   .                      ||||| |.|| :
Db   84 PVTEPQQCCPKCVEFHTPSGLRAPPKSCQHNGTMYQHGEIFSAHELFPSRLPNQCVLCSC 143

Qy   41 PFCHLYWGCTRTGCYGCLAPFCELNLGDKCLDGVLNNSYESD 83
               |||   ||||| |||    :   |
Db  144 TEGQIYCGLTCPEPGCPAP--LPLPDSCCQACKDEASEQSD 183

RESULT 13
US-08-191-866D-81
; Sequence 81, Application US/08191866D
; Patent No. 5783195
; GENERAL INFORMATION:
;   APPLICANT: Cochran, Mark D
;   APPLICANT: Macdonald, Richard D.
;   TITLE OF INVENTION: Recombinant Infectious Bovine
;   TITLE OF INVENTION: Rhinotracheitis Virus S-IBR-052 And Uses Thereof
;   NUMBER OF SEQUENCES: 99
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: John P. White
;     STREET: 1185 Avenue of the Americas
;     CITY: New York
;     STATE: New York
;     COUNTRY: USA
;     ZIP: 10036
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/191,866D
;     FILING DATE: 4 February 1994
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: White, John P.
;     REGISTRATION NUMBER: 28,678
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (212) 278-0400
;     TELEFAX: (212) 391-0525
;     TELEX: 422523
; INFORMATION FOR SEQ ID NO: 81:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 572 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-191-866D-81

Query Match          7.2%;    Score 68;  DB 1;  Length 572;
Best Local Similarity  20.4%;    Pred. No. 32;
Matches   31;  Conservative  14;  Mismatches  41;  Indels  66;  Gaps   5;

Qy   49 CRRTGCYGCLAPFC-----------------ELNLGDKCLDGVLNNSYESDILKN----- 87
               :|| |::|                          : .|||  |.|  |:|
Db  350 CNTTGCPGKTQRDCNQASYSPWFSNRRMVNSIIVVDKGIDATFSLRVWTIPMSQNYWGSE 409

Qy   88 --------YLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFREL 137
                        |:| ||  :|:  ||||| | ||| ||| |||
Db  410 GRLLLGDRIYIYTRSTSWHS------KLQLGVIDISDY----------------- 442

Qy  138 TYQYRQNIPASELPVAVTSRP---DCYWGRNC 166
               |||    ::         |
Db  443 -----NNIRINWTWHNVPSRPGNDECPWGHSC 469

RESULT 14
US-08-185-949B-81
; Sequence 81, Application US/08185949B
; Patent No. 5874279
; GENERAL INFORMATION:
;   APPLICANT: Mark D. Cochran
;   APPLICANT: Richard D. Macdonald
;   TITLE OF INVENTION: Recombinant Infectious Bovine
;   TITLE OF INVENTION: Rhinotracheitis Virus
;   NUMBER OF SEQUENCES: 104
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: John P. White
;     STREET: 1185 Avenue of the Americas
;     CITY: New York
;     STATE: New York
;     COUNTRY: USA
;     ZIP: 10036
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM 330 466 DX2
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/185,949B
;     FILING DATE: 03-NOV-1994
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: White, John P.
;     REGISTRATION NUMBER: ,678
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (212) 278-0400
;     TELEFAX: (212) 278-0525
; INFORMATION FOR SEQ ID NO: 81:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 572 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-08-185-949B-81

Query Match          7.2%;    Score 68;  DB 2;  Length 572;
Best Local Similarity  20.4%;    Pred. No. 32;
Matches   31;  Conservative  14;  Mismatches  41;  Indels  66;  Gaps   5;

Qy   49 CRRTGCYGCLAPFC-----------------ELNLGDKCLDGVLNNSYESDILKN----- 87
               :|| |::|                          : .|||  |.|  |:|
Db  350 CNTTGCPGKTQRDCNQASYSPWFSNRRMVNSIIVVDKGIDATFSLRVWTIPMSQNYWGSE 409

Qy   88 --------YLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFREL 137
                        |:| ||  :|:  ||||| | ||| ||| |||
Db  410 GRLLLGDRIYIYTRSTSWHS------KLQLGVIDISDY----------------- 442

Qy  138 TYQYRQNIPASELPVAVTSRP---DCYWGRNC 166
               |||    ::         |
Db  443 -----NNIRINWTWHNVPSRPGNDECPWGHSC 469
```

Db   904 CHVKGSHSAVCHLETGLCDC----------KPNVTGQQCDQCL---HGYYGL--D 943

Qy   53 GCYGCLAPFCEL--NLGDKCLD 72
                :|:|||| : :|:|||
Db   944 SGHGCRPCNCSVAGSVSDGCTD 965

RESULT 8
US-09-543-681A-5344
; Sequence 5344, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
;   APPLICANT: GARY BRETON
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
;   TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: 2709.1002-001
;   CURRENT APPLICATION NUMBER: US/09/543,681A
;   CURRENT FILING DATE: 2000-04-05
;   PRIOR APPLICATION NUMBER: US 60/128,706
;   PRIOR FILING DATE: 1999-04-09
;   NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5344
;   LENGTH: 200
;   TYPE: PRT
;   ORGANISM: Proteus mirabilis
US-09-543-681A-5344

Query Match   7.5%;   Score 71; DB 4; Length 200;
Best Local Similarity 25.3%;   Pred. No. 3.6;
Matches   20; Conservative 17; Mismatches 34; Indels 8; Gaps 3;

Qy   76 NNNSYESDILKNYLATRGLTWKNMLTE--SLVALQRGVFLLSD-YRVTGDTVLC-----Y 127
                : | ::||| ::|||:| |||:| :: || :|: :
Db   5 SKNPMYQDLIRGELTEAADTLSRFLQDDANIEAIQKAAVLLADSFKAGGKVLSCGNGGSH 64

Qy   128 CCGLRSFRELTYQYRONIP 146
                |||| :|| ||||||
Db   65 CDAMHFAEELTGRYRENRP 83

RESULT 9
US-09-328-352-5062
; Sequence 5062, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;   APPLICANT: Gary L. Breton et al.
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;   TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: GTC99-03PA
;   CURRENT APPLICATION NUMBER: US/09/328,352
;   CURRENT FILING DATE: 1999-06-04
;   NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5062
;   LENGTH: 460
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-5062

Query Match   7.5%;   Score 71; DB 4; Length 460;
Best Local Similarity 26.1%;   Pred. No. 11;
Matches   37; Conservative 22; Mismatches 57; Indels 26; Gaps 9;

Qy   26 QDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELNLGDKCLDGVLNN---NSYES 82
               | || :| :|| :|||| |: |:|| ||||||||| :|:||| :|:| :
Db   333 QSPK-STQPDVICRTNFKYMYWSMACQLTHHTIGG-CNVQGDLMGSGTISGSSTPDSYGS 390

Qy   83 DILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVL---CYCCGLR-SFREL 137
               ||:| ||||:||| || :| :| || |:| ::   : | :|||
Db   391 LLELTWNTTKPLTLANGET------RG--FLQD----GDTLIMKGHCEKNGIRIGFGEV 437

Qy   138 TYQYRQNIPASELPVAVTSRPD 159
               || |:| :|
Db   438 R--NTVLPALTFDFAETSEPN 456

RESULT 10
US-09-845-583A-2
; Sequence 2, Application US/09845583A
; Patent No. 6635616
; GENERAL INFORMATION:
;   APPLICANT: Burgeson, Robert
;   APPLICANT: Brunken, William Joseph
;   APPLICANT: Champliaud, Marie-France
;   APPLICANT: Hunter, Dale
;   TITLE OF INVENTION: LAMININ 15 AND USES THEREOF
;   FILE REFERENCE: 10287-056001
;   CURRENT APPLICATION NUMBER: US/09/845,583A
;   CURRENT FILING DATE: 2001-04-30
;   PRIOR APPLICATION NUMBER: US 60/200,863
;   PRIOR FILING DATE: 2000-05-01
;   NUMBER OF SEQ ID NOS: 18
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 3635
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-845-583A-2

Query Match   7.5%;   Score 71; DB 4; Length 3635;
Best Local Similarity 24.5%;   Pred. No. 1.7e+02;
Matches   25; Conservative 9; Mismatches 38; Indels 30; Gaps 6;

Qy   1 CPLQGSHAL----------------CTCCFQPMPDRRABREQDPRVAPQQCAVCLQP 41
              || ||: | : |:
Db   1379 CPCRG-HVIGRDCSRCATGYWGFNCRPCDCGARLCDЕLTGQCICPRTVPPDCLVC-QP 1436

Qy   42 ---FCHLYWGCTRTGCYG-----CLAPFCELNLGD-KCLDGV 74
                ||||:|| | :|: :   ::| | |: |||||
Db   1437 QSFGCHPLVGCEECNCSGPGVQELTDFTCDMDSGQCRCRPNV 1478

RESULT 11
US-09-798-051-6
; Sequence 6, Application US/09798051
; Patent No. 6632780
; GENERAL INFORMATION:
;   APPLICANT: Zhang, Ke
;   APPLICANT: Cam, Linh
;   APPLICANT: Nakayama, Naoki
;   TITLE OF INVENTION: Chordin-Like-2 Molecules and Uses Thereof
;   FILE REFERENCE: 01-005
;   CURRENT APPLICATION NUMBER: US/09/798,051
;   CURRENT FILING DATE: 2001-03-05
;   NUMBER OF SEQ ID NOS: 21
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 408
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-798-051-6

Query Match   7.2%;   Score 68; DB 4; Length 408;
Best Local Similarity 24.3%;   Pred. No. 20;
Matches   25; Conservative 8; Mismatches 46; Indels 24; Gaps 3;

Qy   2 PLQGSHALCTCCFQP------MPDRRAE-----------REQDPRVAPQQCAVCLQ 40
              | :| :|||||:| |:|: ||:| | :| :|||
Db   63 PVTEPQQCCPKCVEPHTРSGLRAPPKSCGHNGTMYQHGBIFSAHELFPSRLPNQCVLCSC 122

Qy   41 PFCHLYWGCTRTGCYGCLAPFCELNLGDKCLDGVLNNNSYESD 83
               | ||:||| |::|| || |||||||:| ||| :|::|:|
Db   123 TBGQIYCGLITTCPPPGCPAP---LPLPDSCCQACKDEASEQSD 162

RESULT 12
US-09-798-051-5

```
;   INFORMATION FOR SEQ ID NO: 15:
,   SEQUENCE CHARACTERISTICS:
,     LENGTH: 1835 amino acids
,     TYPE: amino acid
,     STRANDEDNESS: not relevant
,     TOPOLOGY: not relevant
,     MOLECULE TYPE: protein
US-08-836-325-15

Query Match         8.0%;  Score 75;  DB 3;  Length 1835;
Best Local Similarity   19.9%;  Pred. No. 24;
Matches   29;  Conservative   26;  Mismatches   59;  Indels   32;  Gaps   6;

Qy     4  QGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCE   63
             : :||:  |:|: |||  : : | : :|: | :  ||||||      :: ||||  :
Db  1023  RSSSSECSTVDNPLPGEGERAEAEPNDEPEACFTD--------GCVRRFC-------CQ 1066

Qy    64  LNLG------DKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSD  116
            :|           : :: ||:  :: : ||: :  :|:  : ::|  :: : :
Db  1067  VNSGKGKWWIRKTCY--IVEHSWFESFIVLMLLSSGA----LAFEDIYIEKKTIKIILE 1120

Qy   117  YRVTGDTVLCYCCGLRSFRBLTYQYR  142
             :|         :
Db  1121  Y--ADKIFTYIFILEMLLKWAYGYK  1143

RESULT 6
US-08-460-309-5
; Sequence 5, Application US/08460309
; Patent No. 5837496
; GENERAL INFORMATION:
;   APPLICANT: Engvall, Eva
;   APPLICANT: Leivo, Ilmo
;   TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
;   TITLE OF INVENTION: Fragments and Uses Thereof
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Campbell and Flores
;     STREET: 4370 La Jolla Village Drive, Suite 700
;     CITY: San Diego
;     STATE: California
;     COUNTRY: USA
;     ZIP: 92122
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/460,309
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/125,077
;     FILING DATE: 22-SEP-1993
;     APPLICATION NUMBER: US PCT/US 94/10730
;     FILING DATE: 21-SEP-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/472,319
;     FILING DATE: 30-JAN-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/919,951
;     FILING DATE: 27-JUL-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Campbell, Cathryn A.
;     REGISTRATION NUMBER: 31,815
;     REFERENCE/DOCKET NUMBER: P-LA 9721
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (619) 535-9001
;     TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 3075 amino acids
```

```
;     TYPE: amino acid
;     TOPOLOGY: linear
US-08-460-309-5

Query Match         7.6%;  Score 72;  DB 2;  Length 3075;
Best Local Similarity   28.0%;  Pred. No. 1e+02;
Matches   23;  Conservative   8;  Mismatches   21;  Indels   30;  Gaps   5;

Qy     1  CPLQGSHA------LCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRT   52
            :|:|||:        |:|:  ||  :  :|
Db   904  CHVKGSHSAVCHLETGLCDC------------KPNVTGQQCDQCL----HGYYGL-D  943

Qy    53  GCYGCLAPFCEL--NLGDKCLD   72
            : ||| ||   :  :: ::|
Db   944  SGHGCRPCNCSVAGSVSDGCTD  965

RESULT 7
US-08-125-077-5
; Sequence 5, Application US/08125077
; Patent No. 5872231
; Patent No. 5872231 5840863
; GENERAL INFORMATION:
;   APPLICANT: Engvall, Eva
;   APPLICANT: Leivo, Ilmo
;   TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
;   TITLE OF INVENTION: Fragments and Uses Thereof
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRSS:
;     ADDRESSBE: Campbell and Flores
;     STREET: 4370 La Jolla Village Drive, Suite 700
;     CITY: San Diego
;     STATE: California
;     COUNTRY: USA
;     ZIP: 92122
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/125,077
;     FILING DATE: 22-SEP-1993
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US PCT/US 94/10730
;     FILING DATE: 21-SEP-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/472,319
;     FILING DATE: 30-JAN-1990
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/919,951
;     FILING DATE: 27-JUL-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Campbell, Cathryn A.
;     REGISTRATION NUMBER: 31,815
;     REFERENCE/DOCKET NUMBER: P-LA 9721
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (619) 535-9001
;     TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 3075 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
US-08-125-077-5

Query Match         7.6%;  Score 72;  DB 2;  Length 3075;
Best Local Similarity   28.0%;  Pred. No. 1e+02;
Matches   23;  Conservative   8;  Mismatches   21;  Indels   30;  Gaps   5;

Qy     1  CPLQGSHA------LCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRT   52
            :|:|||:        |:|:  |||   :::  |:
```

; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12650

Query Match　　8.5%; Score 80; DB 4; Length 197;
Best Local Similarity　　30.6%; Pred. No. 0.34;
Matches 22; Conservative 15; Mismatches 27; Indels 8; Gaps 3;

```
Qy   83 DILKNYLATRGLTWKNMLTE--SLVALQRGVFLLSD-YRVTGDTVLC-----YCCGLRSF 134
           | ::::||    :   ||    :: ::||   :  |:|
Db    8 DLIRNELNEAAETLANFLQDEANIHAIQRAAVLLADSFKAGGKVLSCGNGGSHCDAMHFA 67

Qy  135 RELTYQYRQNIP 146
           ||:|:||
Db   68 EELTGRYRENRP 79
```

RESULT 3
US-08-841-483-2
; Sequence 2, Application US/08841483B
; Patent No. 5976875
; GENERAL INFORMATION:
; APPLICANT: Prescott, Steven M.
; APPLICANT: Bunting, Michaeline
; APPLICANT: Tang, Wen
; APPLICANT: Topham, Matthew
; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
; TITLE OF INVENTION: Methods of Use Thereof
; FILE REFERENCE: 2037.2.1a
; CURRENT APPLICATION NUMBER: US/08/841,483B
; CURRENT FILING DATE: 1997-04-22
; EARLIER APPLICATION NUMBER: 60/016,210
; EARLIER FILING DATE: 1996-04-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-841-483-2

Query Match　　8.1%; Score 76; DB 2; Length 567;
Best Local Similarity　　21.6%; Pred. No. 3.9;
Matches 36; Conservative 21; Mismatches 54; Indels 56; Gaps 10;

```
Qy    3 LQGSHALCTCCFQPMPD---RRAER--------EQDPRVAPQQ----------CAVCL 39
Db   82 LQG--AFCDCCGLRVDEGCLRKADKRFQCKEIMLKNDTKVLDAMPHHWIRGNVPLCSYCM 139

Qy   40 QPFCHLYWGCTRTGC-YGCLAPFCELNLGDKCLDGVLNNSYESDILKN-----YLAT- 91
Db  140 --VCKQQCGCQPKLCDYRCI--WCQKTVHDECMKNSLKNEKCDFGEFKNLIIPPSYLTSI 195

Qy   92 ------------RGLTWKNM--LTESLVALQRGVFLLSDYRV 119
Db  196 NQMRKDKKTDYEVLASKLGKQWTPLIILANSRSGTNMGEGLLGEFRI 242
```

RESULT 4
US-09-382-911-2
; Sequence 2, Application US/09382911
; Patent No. 6221658
; GENERAL INFORMATION:
; APPLICANT: Prescott, Steven M.
; APPLICANT: Bunting, Michaeline
; APPLICANT: Tang, Wen
; APPLICANT: Topham, Matthew
; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
; TITLE OF INVENTION: Methods of Use Thereof
; FILE REFERENCE: 2037.2.1a
; CURRENT APPLICATION NUMBER: US/09/382,911
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 08/841,483
; PRIOR FILING DATE: 1997-04-22

; PRIOR APPLICATION NUMBER: 60/016,210
; PRIOR FILING DATE: 1996-04-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-382-911-2

Query Match　　8.1%; Score 76; DB 3; Length 567;
Best Local Similarity　　21.6%; Pred. No. 3.9;
Matches 36; Conservative 21; Mismatches 54; Indels 56; Gaps 10;

```
Qy    3 LQGSHALCTCCFQPMPD---RRAER--------EQDPRVAPQQ----------CAVCL 39
Db   82 LQG--AFCDCCGLRVDEGCLRKADKRFQCKEIMLKNDTKVLDAMPHHWIRGNVPLCSYCM 139

Qy   40 QPFCHLYWGCTRTGC-YGCLAPFCELNLGDKCLDGVLNNSYESDILKN-----YLAT- 91
Db  140 --VCKQQCGCQPKLCDYRCI--WCQKTVHDECMKNSLKNEKCDFGEFKNLIIPPSYLTSI 195

Qy   92 ------------RGLTWKNM--LTESLVALQRGVFLLSDYRV 119
Db  196 NQMRKDKKTDYEVLASKLGKQWTPLIILANSRSGTNMGEGLLGEFRI 242
```

RESULT 5
US-08-836-325-15
; Sequence 15, Application US/08836325
; Patent No. 6110672
; GENERAL INFORMATION:
; APPLICANT: Mandel, Gail
; APPLICANT: Halegoua, Simon
; APPLICANT: Borden, Laurence A.
; TITLE OF INVENTION: Peripheral Nervous System Specific
; TITLE OF INVENTION: Sodium Channels, DNA Encoding Therefor, Crystallization,
; TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational
; TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Using
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
; STREET: 1100 New York Ave., N. W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,325
; FILING DATE: 2-MAY-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14251
; FILING DATE: 02-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/482,401
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/334,029
; FILING DATE: 02-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0917.0240002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   May 7, 2004, 14:42:23 ; Search time 11.0433 Seconds
                                  (without alignments)
                                  776.028 Million cell updates/sec

Title:          US-10-048-046-2_COPY_476_641
Perfect score:  943
Sequence:       1 CPLQGSHALCTCCFQPMPDR.........ASELPVAVTSRPDCYWGRNC 166

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:     389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued Patents AA:*
            1: /cgn2_6/ptcdata/2/iaa/5A_COMB.pep:*
            2: /cgn2_6/ptcdata/2/iaa/5B_COMB.pep:*
            3: /cgn2_6/ptcdata/2/iaa/6A_COMB.pep:*
            4: /cgn2_6/ptcdata/2/iaa/6B_COMB.pep:*
            5: /cgn2_6/ptcdata/2/iaa/PCTUS_COMB.pep:*
            6: /cgn2_6/ptcdata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                            SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	81	8.6	376	4	US-09-489-039A-11343	Sequence 11343, A
2	80	8.5	197	4	US-09-489-039A-12650	Sequence 12650, A
3	76	8.1	567	2	US-08-841-483-2	Sequence 2, Appli
4	76	8.1	567	3	US-09-382-911-2	Sequence 2, Appli
5	75	8.0	1835	3	US-08-836-325-15	Sequence 15, Appl
6	72	7.6	3075	2	US-08-460-309-5	Sequence 5, Appli
7	72	7.6	3075	2	US-08-125-077-5	Sequence 5, Appli
8	71	7.5	200	4	US-09-543-681A-5344	Sequence 5344, Ap
9	71	7.5	460	4	US-09-328-352-5062	Sequence 5062, Ap
10	71	7.5	3635	4	US-09-845-583A-2	Sequence 2, Appli
11	68	7.2	408	4	US-09-798-051-6	Sequence 6, Appli
12	68	7.2	429	4	US-09-798-051-5	Sequence 5, Appli
13	68	7.2	572	1	US-08-191-866D-81	Sequence 81, Appl
14	68	7.2	572	2	US-08-18-949B-81	Sequence 81, Appl
15	67.5	7.2	452	4	US-09-798-051-9	Sequence 9, Appli
16	67	7.1	174	3	US-09-383-586-12	Sequence 12, Appl
17	67	7.1	202	4	US-09-489-039A-14068	Sequence 14068, A
18	67	7.1	276	2	US-08-786-606-5	Sequence 5, Appli
19	67	7.1	276	4	US-09-205-258-1027	Sequence 1027, Ap
20	67	7.1	341	4	US-09-205-258-1034	Sequence 1034, Ap
21	67	7.1	441	4	US-08-914-375C-66	Sequence 66, Appl
22	67	7.1	449	3	US-09-306-593-10	Sequence 10, Appl
23	67	7.1	503	3	US-09-382-256-10	Sequence 10, Appl
24	67	7.1	503	4	US-09-395-115-10	Sequence 10, Appl
25	67	7.1	503	4	US-08-436-265-10	Sequence 10, Appl
26	67	7.1	503	4	US-09-679-187-10	Sequence 10, Appl
27	67	7.1	3174	2	US-08-477-451-3	Sequence 3, Appli
28	66.5	7.1	259	4	US-09-252-991A-29086	Sequence 29086, A
29	66.5	7.1	387	1	US-08-123-161A-12	Sequence 12, Appl
30	66.5	7.1	387	1	US-08-483-278-12	Sequence 12, Appl
31	66	7.0	125	4	US-09-489-039A-10681	Sequence 10681, A
32	66	7.0	330	4	US-09-489-039A-8234	Sequence 8234, Ap
33	66	7.0	515	4	US-09-635-872A-6	Sequence 6, Appli
34	66	7.0	515	4	US-09-636-077A-6	Sequence 6, Appli
35	66	7.0	515	4	US-09-636-060C-6	Sequence 6, Appli
36	66	7.0	515	4	US-09-098-552-6	Sequence 6, Appli
37	65.5	6.9	276	4	US-09-719-402A-6	Sequence 6, Appli
38	65.5	6.9	324	4	US-09-198-452A-782	Sequence 782, App
39	65.5	6.9	539	4	US-09-719-402A-2	Sequence 2, Appli
40	65.5	6.9	1417	4	US-08-900-230-3	Sequence 3, Appli
41	65	6.9	177	4	US-09-252-991A-28511	Sequence 28511, A
42	65	6.9	212	4	US-09-107-532A-5215	Sequence 5215, Ap
43	65	6.9	1015	1	US-08-537-210A-1	Sequence 1, Appli
44	65	6.9	1015	3	US-09-113-825-1	Sequence 1, Appli
45	65	6.9	1165	1	US-08-144-121-2	Sequence 2, Appli

```
                            ALIGNMENTS

RESULT 1
US-09-489-039A-11343
;  Sequence 11343, Application US/09489039A
;  Patent No. 6610836
;  GENERAL INFORMATION:
;  APPLICANT: Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:   2000-01-27
;  PRIOR APPLICATION NUMBER: US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS: 14342
;  SEQ ID NO 11343
;  LENGTH: 376
;  TYPE: PRT
;  ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11343

  Query Match       8.6%;  Score 81;  DB 4;  Length 376;
  Best Local Similarity  33.7%;  Pred. No. 0.62;
  Matches   30;  Conservative   9;  Mismatches  32;  Indels  18;  Gaps  5;

Qy      32 PQQCAVCL-------QPFCHLYWGCTRTGCYGCLAPFCELNLGDKCLDGVLNNNSYES 82
                        : |:|  |   :|: ||:: .:||   :|::::.|
Db     237 PIAAAACTADAVSSTAGKPLKQLFW-C--AGSWGTLYPFSGNQNGGK---GVIRDSSLLS 290

Qy      83 DILKNYLATRGLTWKNMLTESLVALQRGV 111
               : : : :.|   |:  .||  :| :| |
Db     291 ARVLAALHRRGLAWKTMGSE--AMCRGV 316


RESULT 2
US-09-489-039A-12650
;  Sequence 12650, Application US/09489039A
;  Patent No. 6610836
;  GENERAL INFORMATION:
;  APPLICANT: Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:   2000-01-27
;  PRIOR APPLICATION NUMBER: US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS: 14342
;  SEQ ID NO 12650
;  LENGTH: 197
;  TYPE: PRT
```

```
Query Match             9.1%;  Score  85.5;  DB 4;  Length 244;
Best Local Similarity  27.4%;  Pred. No. 3.3;
Matches    37;  Conservative   11;  Mismatches    42;  Indels    45;  Gaps    10;

Qy      3 LQSHALCTCC------FQPMPDRR----AEREQDPRVAPQQCAVCLQPFCHLYW---- 47
            :|||---|   |||--- :|:|   |:|  ||||: :|||    |||||
Db     17 VKGS---ANCCRPMLRNFRLQPQKRPRKRHVRRELTPSV--RWCAATFQPFCWLAWCRLS 71

Qy     48 ------GCTRTGCY----GCLAPFCEL---NLGDKCLDGVLNNNSYESDILKNYLAT-- 91
                  |  :||| |---|||:||| |:|  ||||||:||||||||| :|  :|
Db     72 PEWCWGRCDDTGCVVRARAC-RPACHRLSTGGLGDR--RRVIHRNETDWRILGAQTTGG 128

Qy     92 ------RGLTWKNML 100
                  |---|| ::
Db    129 VRHTPHRGREWPGLI 143
```

Query Match   14.6%;   Score 137.5;   DB 3;   Length 453;
Best Local Similarity 25.0%;   Pred. No. 4.9e-05;
Matches   45;   Conservative 26;   Mismatches 88;   Indels 21;   Gaps 6;

```
Qy    1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP  60
         |:|||    ||:||   :||| |
Db  255 CEHHGAHLQCHLCQGMMPFRANLQ-----VPLHCKGCDRPFCGAYWS-SENVTQGVSGP  307

Qy   61 FCELN----IGDKCLDGV---LNNNSYESDILKNYLATRGLTWKNMLTESL----VAL  107
         :| |      : :|       :|    :|
Db  308 VCVRETFRPISERTITRIPFITHEMNRHEQDITQRCIAHMEKTVPDVVAEWLRLFNNREI  367

Qy  108 QRGVFLLSDYR-VTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC  166
          :|   :  : |:  : |  :|:   : :  : :| : |:   |  ||  :|||:|:| |
Db  368 DRSRMPLNHAETITASTHVCNDCYDKLVGFLLYWFRITLPRNHLPADVVAREDCWYGYAC  427
```

RESULT 14
AAB34759
ID   AAB34759 standard; protein; 227 AA.
XX
AC   AAB34759;
XX
DT   26-JAN-2001 (first entry)
XX
DE   Human secreted protein fragment encoded by DNA clone vo3 1.
XX
KW   Secreted protein; human; autoimmune disorder; multiple sclerosis; ulcer;
KW   systemic lupus erythematosus; rheumatoid arthritis; anaemia; stroke;
KW   haematopoiesis regulation; tissue regrowth; wound healing; haemophilia;
KW   Alzheimer's disease; Parkinson's disease; Shy-drager syndrome; cancer;
KW   contraceptive; infection; growth inhibition; hyperproliferative disorder;
KW   psoriasis.
XX
OS   Homo sapiens.
XX
XX   WO200055375-A1.
PN   WO200055375-A1.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US007285.
XX
PR   17-MAR-1999;   99US-0124808P.
PR   17-MAR-1999;   99US-0124916P.
PR   17-AUG-1999;   99US-0149639P.
PR   01-OCT-1999;   99US-0157247P.
PR   29-NOV-1999;   99US-0167824P.
PR   15-FEB-2000; 2000US-0182711P.
XX
PA   (ALPH-) ALPHAGENE INC.
XX
PI   Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;

XX
DR   WPI; 2000-638211/61.
DR   N-PSDB; AAC59810.
XX
PT   Novel proteins and polypeptides useful for the treatment of e.g multiple
PT   sclerosis, systemic lupus erythematosus, rheumatoid arthritis, cancer,
PT   Alzheimer's disease, Parkinson's disease, stroke, anemia and ulcers.
XX
PS   Disclosure; Page 480-481; 493pp; English.
XX
CC   This invention relates to 59 human secreted proteins and the nucleotide
CC   sequences encoding them. Sequences AAC59788-C59846 and AAB34687-B34745
CC   represent the proteins and their encoding nucleotide sequences, and
CC   sequences AAB34746-B34771 represent fragments of the proteins. Probes for
CC   the DNA sequences are represented by sequences AAC59847-C59596. The
CC   proteins exhibit neuroprotective, dermatological, immunosuppressive,
CC   antiinflammatory, antianaemic, nootropic, antiparkinsonian,
CC   cerebroprotective, haemostatic, vulnerary, cytostatic, antipsoriatic,
CC   antibacterial, virucide, and fungicide activity. The proteins and
CC   nucleotide sequences are useful as nutritional sources or supplements and
CC   in research. The proteins are useful for treating immune deficiency and
CC   disorders, which may be genetic or resulting from infections, autoimmune
CC   disorders such as multiple sclerosis, systemic lupus erythematosus,
CC   rheumatoid arthritis, and for treating myeloid or lymphoid cell
CC   deficiencies such as anaemias by regulating haematopoiesis. The proteins
CC   are also useful in compositions for bone, cartilage, tendon, ligament
CC   and/or nerve tissue growth or regeneration, for wound healing, tissue
CC   repair and replacement and in the treatment of wounds, incisions and
CC   ulcers. Other uses include in the treatment of central and peripheral
CC   nervous system and neuropathies such as Alzheimer's and Parkinson's
CC   diseases and Shy-Drager syndrome, and mechanical and traumatic disorders,
CC   such as spinal cord disorders, head trauma and stroke. The proteins may
CC   also be used as a contraceptive, and for treating coagulation disorders
CC   such as haemophilias. The protein and nucleotide sequences with cadherin
CC   activity are useful for treating cancer. Other uses for the protein
CC   include for inhibiting the growth, infection or function of, or killing,
CC   infectious agents such as bacteria, virus, fungi and other parasites, for
CC   effecting bodily characteristics such as height, weight, hair colour,
CC   effecting biorhythms or cardiac cycles or rhythms, effecting metabolism,
CC   catabolism, anabolism, processing, utilization, storage or elimination of
CC   dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors,
CC   effecting behavioural characteristics, providing analgesic effects and
CC   for treating hyperproliferative disorders such as psoriasis
XX
SQ   Sequence 227 AA;

Query Match   9.8%;   Score 92.5;   DB 3;   Length 227;
Best Local Similarity 23.8%;   Pred. No. 0.62;
Matches   45;   Conservative 21;   Mismatches 70;   Indels 53;   Gaps 10;

```
Qy    2 PLQGSHALCTCCFQPMP---DRRAEREQDPRVAP------QQCAVC-------------  38
         |:||||| || |:|      :||  ||:
Db   19 PVLPSEAL--LFPGLPAGFSRLSSNAGPRLLAWVLACPLRPLAACLLSLVALPGCWAA  75

Qy   39 ----LQPFCHLYWGCTRTG--CYGCLAP-FCELNLGDKCLDGVLNNNSYESDILKNYLAT  91
             :||   :  :|      || :||     :|:
Db   76 LSGRLLPVCFPWWLCLGAGPAFSGCLLPVYCHLQRGSLLRPTLLH------LAPPWL--  126

Qy   92 RGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVL---CYCCGLRSFRELTYQYRQNIPAS  148
          :|      ::||:|  |||    :|:  |:|
Db  127 --LAWPNLAFCAMLELELLLFFRGGNRVESGKGLAPKCCCCGFFAFS------KDALPGP  178

Qy  149 ELPVAVTSR  157
          ||:||:|
Db  179 KLQTAVLSK  187
```

Sequence 227 AA;

RESULT 15
AAU56425
ID   AAU56425 standard; protein; 244 AA.
XX
AC   AAU56425;
XX

```
PR  28-APR-1999;  99US-0131449P.          PR  21-JUL-1999;  99US-0144814P.
PR  30-APR-1999;  99US-0132048P.          PR  21-JUL-1999;  99US-0145086P.
PR  30-APR-1999;  99US-0132407P.          PR  21-JUL-1999;  99US-0145088P.
PR  04-MAY-1999;  99US-0132484P.          PR  22-JUL-1999;  99US-0145085P.
PR  05-MAY-1999;  99US-0132485P.          PR  22-JUL-1999;  99US-0145087P.
PR  06-MAY-1999;  99US-0132486P.          PR  22-JUL-1999;  99US-0145089P.
PR  06-MAY-1999;  99US-0132487P.          PR  22-JUL-1999;  99US-0145192P.
PR  07-MAY-1999;  99US-0132863P.          PR  23-JUL-1999;  99US-0145145P.
PR  11-MAY-1999;  99US-0134256P.          PR  23-JUL-1999;  99US-0145218P.
PR  14-MAY-1999;  99US-0134218P.          PR  23-JUL-1999;  99US-0145224P.
PR  14-MAY-1999;  99US-0134219P.          PR  26-JUL-1999;  99US-0145276P.
PR  14-MAY-1999;  99US-0134221P.          PR  27-JUL-1999;  99US-0145913P.
PR  18-MAY-1999;  99US-0134370P.          PR  27-JUL-1999;  99US-0145918P.
PR  19-MAY-1999;  99US-0134768P.          PR  27-JUL-1999;  99US-0145919P.
PR  20-MAY-1999;  99US-0134941P.          PR  28-JUL-1999;  99US-0145951P.
PR  21-MAY-1999;  99US-0135124P.          PR  02-AUG-1999;  99US-0146386P.
PR  24-MAY-1999;  99US-0135353P.          PR  02-AUG-1999;  99US-0146388P.
PR  25-MAY-1999;  99US-0135629P.          PR  03-AUG-1999;  99US-0146389P.
PR  27-MAY-1999;  99US-0136021P.          PR  04-AUG-1999;  99US-0147038P.
PR  28-MAY-1999;  99US-0136392P.          PR  04-AUG-1999;  99US-0147204P.
PR  01-JUN-1999;  99US-0136782P.          PR  05-AUG-1999;  99US-0147302P.
PR  03-JUN-1999;  99US-0137222P.          PR  05-AUG-1999;  99US-0147192P.
PR  04-JUN-1999;  99US-0137528P.          PR  06-AUG-1999;  99US-0147260P.
PR  07-JUN-1999;  99US-0137502P.          PR  06-AUG-1999;  99US-0147303P.
PR  08-JUN-1999;  99US-0137724P.          PR  09-AUG-1999;  99US-0147416P.
PR  10-JUN-1999;  99US-0138094P.          PR  09-AUG-1999;  99US-0147493P.
PR  10-JUN-1999;  99US-0138540P.          PR  10-AUG-1999;  99US-0147935P.
PR  14-JUN-1999;  99US-0138847P.          PR  11-AUG-1999;  99US-0148171P.
PR  16-JUN-1999;  99US-0139119P.          PR  12-AUG-1999;  99US-0148319P.
PR  16-JUN-1999;  99US-0139452P.          PR  13-AUG-1999;  99US-0148341P.
PR  17-JUN-1999;  99US-0139453P.          PR  13-AUG-1999;  99US-0148565P.
PR  18-JUN-1999;  99US-0139492P.          PR  16-AUG-1999;  99US-0148684P.
PR  18-JUN-1999;  99US-0139454P.          PR  17-AUG-1999;  99US-0149368P.
PR  18-JUN-1999;  99US-0139455P.          PR  18-AUG-1999;  99US-0149175P.
PR  18-JUN-1999;  99US-0139456P.          PR  20-AUG-1999;  99US-0149426P.
PR  18-JUN-1999;  99US-0139457P.          PR  20-AUG-1999;  99US-0149722P.
PR  18-JUN-1999;  99US-0139458P.          PR  20-AUG-1999;  99US-0149723P.
PR  18-JUN-1999;  99US-0139459P.          PR  23-AUG-1999;  99US-0149929P.
PR  18-JUN-1999;  99US-0139460P.          PR  25-AUG-1999;  99US-0149902P.
PR  18-JUN-1999;  99US-0139461P.          PR  26-AUG-1999;  99US-0149930P.
PR  18-JUN-1999;  99US-0139462P.          PR  27-AUG-1999;  99US-0150566P.
PR  18-JUN-1999;  99US-0139463P.          PR  27-AUG-1999;  99US-0150884P.
PR  18-JUN-1999;  99US-0139750P.          PR  30-AUG-1999;  99US-0151065P.
PR  18-JUN-1999;  99US-0139763P.          PR  31-AUG-1999;  99US-0151066P.
PR  21-JUN-1999;  99US-0139817P.          PR  01-SEP-1999;  99US-0151080P.
PR  22-JUN-1999;  99US-0139899P.          PR  07-SEP-1999;  99US-0151303P.
PR  23-JUN-1999;  99US-0140353P.          PR  10-SEP-1999;  99US-0151438P.
PR  23-JUN-1999;  99US-0140354P.          PR  13-SEP-1999;  99US-0151930P.
PR  24-JUN-1999;  99US-0140695P.          PR  15-SEP-1999;  99US-0152363P.
PR  28-JUN-1999;  99US-0140823P.          PR  16-SEP-1999;  99US-0153070P.
PR  29-JUN-1999;  99US-0140991P.          PR  20-SEP-1999;  99US-0153758P.
PR  30-JUN-1999;  99US-0141287P.          PR  22-SEP-1999;  99US-0154018P.
PR  01-JUL-1999;  99US-0141842P.          PR  23-SEP-1999;  99US-0154039P.
PR  01-JUL-1999;  99US-0142154P.          PR  24-SEP-1999;  99US-0154779P.
PR  02-JUL-1999;  99US-0142055P.          PR  28-SEP-1999;  99US-0155139P.
PR  06-JUL-1999;  99US-0142390P.          PR  29-SEP-1999;  99US-0155486P.
PR  08-JUL-1999;  99US-0142803P.          PR  04-OCT-1999;  99US-0155659P.
PR  09-JUL-1999;  99US-0142920P.          PR  05-OCT-1999;  99US-0156458P.
PR  12-JUL-1999;  99US-0142977P.          PR  06-OCT-1999;  99US-0156596P.
PR  13-JUL-1999;  99US-0143542P.          PR  07-OCT-1999;  99US-0157117P.
PR  14-JUL-1999;  99US-0144086P.          PR  08-OCT-1999;  99US-0157753P.
PR  15-JUL-1999;  99US-0144005P.          PR  12-OCT-1999;  99US-0157865P.
PR  16-JUL-1999;  99US-0144085P.          PR  13-OCT-1999;  99US-0158029P.
PR  19-JUL-1999;  99US-0144325P.          PR  13-OCT-1999;  99US-0158232P.
PR  19-JUL-1999;  99US-0144331P.          PR  13-OCT-1999;  99US-0158369P.
PR  19-JUL-1999;  99US-0144332P.          PR  14-OCT-1999;  99US-0159293P.
PR  19-JUL-1999;  99US-0144333P.          PR  14-OCT-1999;  99US-0159294P.
PR  19-JUL-1999;  99US-0144334P.          PR  14-OCT-1999;  99US-0159295P.
PR  20-JUL-1999;  99US-0144335P.          PR  14-OCT-1999;  99US-0159329P.
PR  20-JUL-1999;  99US-0144352P.          PR  14-OCT-1999;  99US-0159330P.
PR  20-JUL-1999;  99US-0144632P.          PR  14-OCT-1999;  99US-0159331P.
PR  20-JUL-1999;  99US-0144884P.          PR  14-OCT-1999;  99US-0159637P.
                                          PR  14-OCT-1999;  99US-0159638P.
```

PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 25-AUG-1999; 99US-0149930P.
PR 26-AUG-1999; 99US-0150566P.
PR 27-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 30-AUG-1999; 99US-0151066P.
PR 31-AUG-1999; 99US-0151080P.
PR 01-SEP-1999; 99US-0151303P.
PR 07-SEP-1999; 99US-0151438P.
PR 10-SEP-1999; 99US-0151930P.
PR 13-SEP-1999; 99US-0152363P.
PR 15-SEP-1999; 99US-0153070P.
PR 16-SEP-1999; 99US-0153758P.
PR 20-SEP-1999; 99US-0154018P.
PR 22-SEP-1999; 99US-0154039P.
PR 23-SEP-1999; 99US-0154779P.
PR 24-SEP-1999; 99US-0155139P.
PR 28-SEP-1999; 99US-0155486P.
PR 29-SEP-1999; 99US-0155659P.
PR 04-OCT-1999; 99US-0155658P.
PR 05-OCT-1999; 99US-0155696P.
PR 06-OCT-1999; 99US-0157117P.
PR 07-OCT-1999; 99US-0157753P.
PR 08-OCT-1999; 99US-0157865P.
PR 12-OCT-1999; 99US-0158029P.
PR 13-OCT-1999; 99US-0158232P.
PR 13-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 14-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 18-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.

PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match   14.6%;   Score 137.5;   DB 3;   Length 453;
Best Local Similarity   25.0%;   Pred. No. 4.9e-05;
Matches   45;   Conservative   26;   Mismatches   88;   Indels   21;   Gaps   6;

```
Qy    1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPPCHLYWGCTRTGCYGCLAP  60
            : |    |    |          |   |          :             |||:      |:   |
Db  255 CEHHGAHLQCHLCQGMPFRANLQ------VPLHCKGCGDRPFCGAYWS-SENVTQGVSGP 307

Qy   61 FCELN----LGDKCLDGV----LNNNSYESDILKNYLATRGLTWKNMLTESL-----VAL 107
              :                 |     ::                  :        ||        :         |:
Db  308 VCVRETFRPISERTITRIPFITHEMNRHEQDITQRCIAHMEKTVPDVVAEWLRLFNNREI 367

Qy  108 QRGVFLLSDYR-VTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
              :   :      :   |:     |   ::          |     :   : :    ||:    : :
Db  368 DRSRMPLNHAEMITASTHVCNDCYDKLVGFLLYWFRITLPRNHLPADVAAREDCWYGYAC 427
```

RESULT 13
AAG37658
ID   AAG37658 standard; protein; 453 AA.
XX
AC   AAG37658;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 46342.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;   2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.

PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 29-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match   14.6%;   Score 137.5; DB 3;   Length 426;
Best Local Similarity 25.0%; Pred. No. 4.6e-05;
Matches 45; Conservative 26; Mismatches 88; Indels 21; Gaps 6;

Qy   1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
             : :| | : | :|:| |: | |
Db   228 CEHHGAHLQCHLCQGMPFRANLQ------VPLHCKGCDRPFCGAYWS-SENVTQGVSGP 280

Qy   61 FCELN----LGDKCLDGV---LNNNSYESDILKNYLATRGLTWKNMLTESL----VAL 107
             : ::|: :| | :|:| : : :| :|
Db   281 VCVRETFRPISERTITRIPFITHEMNRHEQDITQRCIAHMEKTVPDVVAEWLRLFNNREI 340

Qy   108 QRGVFLLSDYR-VTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
             : :   :| :|: ||:|: :: |: | || :|:| : :
Db   341 DRSRMPLNHAETITASTHVCNDCYDKLVGFLLYWFRITLPRNHLPADVAARBDCWYGYAC 400

RESULT 12
AAG20352
ID AAG20352 standard; protein; 453 AA.
XX
AC AAG20352;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 22505.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0163392P.
PR 28-MAY-1999; 99US-0163782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.

```
PR  30-APR-1999;  99US-0132407P.
PR  04-MAY-1999;  99US-0132484P.
PR  05-MAY-1999;  99US-0132485P.
PR  06-MAY-1999;  99US-0132486P.
PR  06-MAY-1999;  99US-0132487P.
PR  07-MAY-1999;  99US-0132863P.
PR  11-MAY-1999;  99US-0134256P.
PR  14-MAY-1999;  99US-0134218P.
PR  14-MAY-1999;  99US-0134219P.
PR  14-MAY-1999;  99US-0134221P.
PR  14-MAY-1999;  99US-0134370P.
PR  18-MAY-1999;  99US-0134768P.
PR  19-MAY-1999;  99US-0134941P.
PR  20-MAY-1999;  99US-0135124P.
PR  21-MAY-1999;  99US-0135353P.
PR  24-MAY-1999;  99US-0135629P.
PR  25-MAY-1999;  99US-0136021P.
PR  27-MAY-1999;  99US-0136392P.
PR  28-MAY-1999;  99US-0136782P.
PR  01-JUN-1999;  99US-0137222P.
PR  03-JUN-1999;  99US-0137528P.
PR  04-JUN-1999;  99US-0137502P.
PR  07-JUN-1999;  99US-0137724P.
PR  08-JUN-1999;  99US-0138094P.
PR  10-JUN-1999;  99US-0138540P.
PR  10-JUN-1999;  99US-0138847P.
PR  14-JUN-1999;  99US-0139119P.
PR  16-JUN-1999;  99US-0139452P.
PR  16-JUN-1999;  99US-0139453P.
PR  17-JUN-1999;  99US-0139492P.
PR  18-JUN-1999;  99US-0139454P.
PR  18-JUN-1999;  99US-0139455P.
PR  18-JUN-1999;  99US-0139456P.
PR  18-JUN-1999;  99US-0139457P.
PR  18-JUN-1999;  99US-0139458P.
PR  18-JUN-1999;  99US-0139459P.
PR  18-JUN-1999;  99US-0139460P.
PR  18-JUN-1999;  99US-0139461P.
PR  18-JUN-1999;  99US-0139462P.
PR  18-JUN-1999;  99US-0139463P.
PR  18-JUN-1999;  99US-0139750P.
PR  21-JUN-1999;  99US-0139763P.
PR  21-JUN-1999;  99US-0139817P.
PR  22-JUN-1999;  99US-0139899P.
PR  23-JUN-1999;  99US-0140353P.
PR  23-JUN-1999;  99US-0140354P.
PR  24-JUN-1999;  99US-0140695P.
PR  28-JUN-1999;  99US-0140823P.
PR  29-JUN-1999;  99US-0140991P.
PR  30-JUN-1999;  99US-0141287P.
PR  01-JUL-1999;  99US-0141842P.
PR  01-JUL-1999;  99US-0142154P.
PR  02-JUL-1999;  99US-0142055P.
PR  06-JUL-1999;  99US-0142390P.
PR  08-JUL-1999;  99US-0142803P.
PR  09-JUL-1999;  99US-0142920P.
PR  12-JUL-1999;  99US-0142977P.
PR  13-JUL-1999;  99US-0143542P.
PR  14-JUL-1999;  99US-0143624P.
PR  15-JUL-1999;  99US-0144005P.
PR  16-JUL-1999;  99US-0144085P.
PR  16-JUL-1999;  99US-0144086P.
PR  19-JUL-1999;  99US-0144325P.
PR  19-JUL-1999;  99US-0144331P.
PR  19-JUL-1999;  99US-0144332P.
PR  19-JUL-1999;  99US-0144333P.
PR  19-JUL-1999;  99US-0144334P.
PR  19-JUL-1999;  99US-0144335P.
PR  20-JUL-1999;  99US-0144352P.
PR  20-JUL-1999;  99US-0144432P.
PR  21-JUL-1999;  99US-0144814P.
PR  21-JUL-1999;  99US-0145086P.

PR  21-JUL-1999;  99US-0145088P.
PR  22-JUL-1999;  99US-0145085P.
PR  22-JUL-1999;  99US-0145087P.
PR  22-JUL-1999;  99US-0145089P.
PR  22-JUL-1999;  99US-0145192P.
PR  23-JUL-1999;  99US-0145145P.
PR  23-JUL-1999;  99US-0145218P.
PR  23-JUL-1999;  99US-0145224P.
PR  26-JUL-1999;  99US-0145276P.
PR  27-JUL-1999;  99US-0145913P.
PR  27-JUL-1999;  99US-0145918P.
PR  27-JUL-1999;  99US-0145919P.
PR  28-JUL-1999;  99US-0145951P.
PR  02-AUG-1999;  99US-0146386P.
PR  02-AUG-1999;  99US-0146388P.
PR  03-AUG-1999;  99US-0146389P.
PR  03-AUG-1999;  99US-0147038P.
PR  04-AUG-1999;  99US-0147204P.
PR  04-AUG-1999;  99US-0147302P.
PR  05-AUG-1999;  99US-0147192P.
PR  05-AUG-1999;  99US-0147260P.
PR  06-AUG-1999;  99US-0147303P.
PR  06-AUG-1999;  99US-0147416P.
PR  09-AUG-1999;  99US-0147493P.
PR  09-AUG-1999;  99US-0147935P.
PR  10-AUG-1999;  99US-0148171P.
PR  11-AUG-1999;  99US-0148319P.
PR  12-AUG-1999;  99US-0148341P.
PR  13-AUG-1999;  99US-0148565P.
PR  13-AUG-1999;  99US-0148684P.
PR  16-AUG-1999;  99US-0149368P.
PR  17-AUG-1999;  99US-0149175P.
PR  18-AUG-1999;  99US-0149426P.
PR  20-AUG-1999;  99US-0149722P.
PR  20-AUG-1999;  99US-0149723P.
PR  20-AUG-1999;  99US-0149929P.
PR  23-AUG-1999;  99US-0149902P.
PR  23-AUG-1999;  99US-0149930P.
PR  25-AUG-1999;  99US-0150566P.
PR  26-AUG-1999;  99US-0150884P.
PR  27-AUG-1999;  99US-0151065P.
PR  27-AUG-1999;  99US-0151066P.
PR  27-AUG-1999;  99US-0151080P.
PR  30-AUG-1999;  99US-0151303P.
PR  31-AUG-1999;  99US-0151438P.
PR  01-SEP-1999;  99US-0151930P.
PR  07-SEP-1999;  99US-0152363P.
PR  10-SEP-1999;  99US-0153070P.
PR  13-SEP-1999;  99US-0153758P.
PR  15-SEP-1999;  99US-0154018P.
PR  16-SEP-1999;  99US-0154039P.
PR  20-SEP-1999;  99US-0154779P.
PR  22-SEP-1999;  99US-0155139P.
PR  23-SEP-1999;  99US-0155486P.
PR  24-SEP-1999;  99US-0155659P.
PR  28-SEP-1999;  99US-0156458P.
PR  29-SEP-1999;  99US-0156596P.
PR  04-OCT-1999;  99US-0157117P.
PR  05-OCT-1999;  99US-0157753P.
PR  06-OCT-1999;  99US-0157865P.
PR  07-OCT-1999;  99US-0158029P.
PR  08-OCT-1999;  99US-0158232P.
PR  12-OCT-1999;  99US-0158369P.
PR  13-OCT-1999;  99US-0159293P.
PR  13-OCT-1999;  99US-0159294P.
PR  13-OCT-1999;  99US-0159295P.
PR  14-OCT-1999;  99US-0159329P.
PR  14-OCT-1999;  99US-0159330P.
PR  14-OCT-1999;  99US-0159331P.
PR  14-OCT-1999;  99US-0159637P.
PR  14-OCT-1999;  99US-0159638P.
PR  18-OCT-1999;  99US-0159584P.
PR  21-OCT-1999;  99US-0160741P.
```

```
Query Match             14.6%;  Score 137.5;  DB 3;  Length 426;
Best Local Similarity   25.0%;  Pred. No. 4.6e-05;
Matches   45;  Conservative  26;  Mismatches  88;  Indels  21;  Gaps  6;

Qy     1  CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP   60
                                            :    : |:||:|
Db   228  CEHHGAHLQCHLCQGMMPFRANLQ------VPLHCKGCDRPFCGAYWS-SENVTQGVSGP  280

Qy    61  FCELN----LGDKCLDGV----LNNNSYESDLLKNYLATRGLTWKNMLTESL----VAL  107
                         :|: :|:     :   :          :|
Db   281  VCVRETFRPISERTITRIPFITHEMNRHEQDITQRCIAHMEKTVPDVVAEWLRLFNNREI  340

Qy   108  QRGVFLLSDYR-VTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC  166
             |   || : |  |:||: :| |        |:|        | |::|  |:
Db   341  DRSRMPLNHAEMITASTHVCNDCYDKLVGFLLYWFRITLPRNHLPADVAAREDCWYGYAC  400
```

RESULT 11
AAG37659
ID   AAG37659 standard; protein; 426 AA.
XX
AC   AAG37659;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 46343.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.

```
PR   22-JUL-1999;  99US-0150085P.
PR   22-JUL-1999;  99US-0150087P.
PR   22-JUL-1999;  99US-0150089P.
PR   22-JUL-1999;  99US-0151192P.
PR   23-JUL-1999;  99US-0151145P.
PR   23-JUL-1999;  99US-0151218P.
PR   23-JUL-1999;  99US-0151224P.
PR   26-JUL-1999;  99US-0151276P.
PR   27-JUL-1999;  99US-0151913P.
PR   27-JUL-1999;  99US-0159918P.
PR   27-JUL-1999;  99US-0159919P.
PR   28-JUL-1999;  99US-0159951P.
PR   02-AUG-1999;  99US-0163386P.
PR   02-AUG-1999;  99US-0163388P.
PR   03-AUG-1999;  99US-0163389P.
PR   03-AUG-1999;  99US-0147038P.
PR   04-AUG-1999;  99US-0147204P.
PR   04-AUG-1999;  99US-0147302P.
PR   05-AUG-1999;  99US-0147192P.
PR   05-AUG-1999;  99US-0147260P.
PR   06-AUG-1999;  99US-0147303P.
PR   06-AUG-1999;  99US-0147416P.
PR   09-AUG-1999;  99US-0147493P.
PR   09-AUG-1999;  99US-0147935P.
PR   10-AUG-1999;  99US-0148171P.
PR   11-AUG-1999;  99US-0148319P.
PR   12-AUG-1999;  99US-0148341P.
PR   13-AUG-1999;  99US-0149565P.
PR   13-AUG-1999;  99US-0148684P.
PR   16-AUG-1999;  99US-0149368P.
PR   17-AUG-1999;  99US-0149175P.
PR   18-AUG-1999;  99US-0149426P.
PR   20-AUG-1999;  99US-0149722P.
PR   20-AUG-1999;  99US-0149723P.
PR   20-AUG-1999;  99US-0149929P.
PR   23-AUG-1999;  99US-0149902P.
PR   23-AUG-1999;  99US-0149930P.
PR   25-AUG-1999;  99US-0150566P.
PR   26-AUG-1999;  99US-0150884P.
PR   27-AUG-1999;  99US-0151065P.
PR   27-AUG-1999;  99US-0151066P.
PR   30-AUG-1999;  99US-0151080P.
PR   31-AUG-1999;  99US-0151303P.
PR   31-AUG-1999;  99US-0151438P.
PR   01-SEP-1999;  99US-0151930P.
PR   07-SEP-1999;  99US-0152363P.
PR   10-SEP-1999;  99US-0153070P.
PR   13-SEP-1999;  99US-0153758P.
PR   15-SEP-1999;  99US-0154018P.
PR   16-SEP-1999;  99US-0154039P.
PR   20-SEP-1999;  99US-0154779P.
PR   23-SEP-1999;  99US-0155139P.
PR   23-SEP-1999;  99US-0155486P.
PR   24-SEP-1999;  99US-0155659P.
PR   28-SEP-1999;  99US-0156458P.
PR   29-SEP-1999;  99US-0156596P.
PR   04-OCT-1999;  99US-0157117P.
PR   05-OCT-1999;  99US-0157753P.
PR   06-OCT-1999;  99US-0157865P.
PR   07-OCT-1999;  99US-0158029P.
PR   08-OCT-1999;  99US-0158232P.
PR   08-OCT-1999;  99US-0158369P.
PR   12-OCT-1999;  99US-0159293P.
PR   13-OCT-1999;  99US-0159294P.
PR   13-OCT-1999;  99US-0159295P.
PR   13-OCT-1999;  99US-0159329P.
PR   14-OCT-1999;  99US-0159329P.
PR   14-OCT-1999;  99US-0159330P.
PR   14-OCT-1999;  99US-0159331P.
PR   14-OCT-1999;  99US-0159637P.
PR   14-OCT-1999;  99US-0159638P.
PR   18-OCT-1999;  99US-0159584P.
PR   21-OCT-1999;  99US-0160741P.
PR   21-OCT-1999;  99US-0160767P.

PR   21-OCT-1999;  99US-0160768P.
PR   21-OCT-1999;  99US-0160770P.
PR   21-OCT-1999;  99US-0160814P.
PR   21-OCT-1999;  99US-0160815P.
PR   22-OCT-1999;  99US-0160980P.
PR   22-OCT-1999;  99US-0160981P.
PR   22-OCT-1999;  99US-0160989P.
PR   22-OCT-1999;  99US-0161404P.
PR   25-OCT-1999;  99US-0161405P.
PR   25-OCT-1999;  99US-0161406P.
PR   25-OCT-1999;  99US-0161359P.
PR   26-OCT-1999;  99US-0161360P.
PR   26-OCT-1999;  99US-0161361P.
PR   28-OCT-1999;  99US-0161920P.
PR   28-OCT-1999;  99US-0161992P.
PR   28-OCT-1999;  99US-0161993P.
PR   29-OCT-1999;  99US-0162142P.

PR   25-FEB-1999;  99US-0121825P.
PR   05-MAR-1999;  99US-0123180P.
PR   09-MAR-1999;  99US-0123588P.
PR   23-MAR-1999;  99US-0125788P.
PR   25-MAR-1999;  99US-0126264P.
PR   29-MAR-1999;  99US-0126785P.
PR   01-APR-1999;  99US-0127462P.
PR   06-APR-1999;  99US-0128234P.
PR   08-APR-1999;  99US-0128714P.
PR   16-APR-1999;  99US-0129845P.
PR   19-APR-1999;  99US-0130077P.
PR   21-APR-1999;  99US-0130449P.
PR   23-APR-1999;  99US-0130510P.
PR   23-APR-1999;  99US-0130891P.
PR   28-APR-1999;  99US-0131449P.
PR   30-APR-1999;  99US-0132048P.
```

```
PR   21-OCT-1999;   99US-0160770P.
PR   21-OCT-1999;   99US-0160814P.
PR   22-OCT-1999;   99US-0160815P.
PR   22-OCT-1999;   99US-0160980P.
PR   22-OCT-1999;   99US-0160981P.
PR   22-OCT-1999;   99US-0160989P.
PR   25-OCT-1999;   99US-0161404P.
PR   25-OCT-1999;   99US-0161405P.
PR   25-OCT-1999;   99US-0161406P.
PR   26-OCT-1999;   99US-0161359P.
PR   26-OCT-1999;   99US-0161360P.
PR   26-OCT-1999;   99US-0161361P.
PR   28-OCT-1999;   99US-0161920P.
PR   28-OCT-1999;   99US-0161992P.
PR   28-OCT-1999;   99US-0161993P.
PR   29-OCT-1999;   99US-0162142P.


Query Match    14.6%;    Score 137.5;   DB 3;   Length 350;
Best Local Similarity   25.0%;    Pred. No. 3.6e-05;
Matches   45;  Conservative  26;  Mismatches  88;  Indels  21;  Gaps  6;

Qy     1   CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP   60
                 ::||       |         :     ||      :
Db   152   CEHHGAHLQCHLCQGGMMPFRANLQ------VPLHCKGCDRPFCGAYWS-SENVTQGVSGP  204

Qy    61   FCELN----LGDKCLDGV---LNNNSYESDILKNYLATRGLTWKNMLTESL-----VAL  107
                         :|:        :|:   :|
Db   205   VCVRETFRPISERTITRIPFITHEMNRHEQDITQRCIAHMEKTVPDVVABWLRLFNNREI  264

Qy   108   QRGVFLLSDYR-VTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC  166
                     :|:    |      :    :   |   |::::|
Db   265   DRSRMPLNHAETITASTHVCNDCYDKLVGFLLYWFRITLPRNHLPADVAAREDCWYGYAC  324


RESULT 10
AAG20353
ID   AAG20353 standard; protein; 426 AA.
XX
AC   AAG20353;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 22506.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;   99US-0121825P.
PR   05-MAR-1999;   99US-0123180P.
PR   09-MAR-1999;   99US-0123548P.
PR   23-MAR-1999;   99US-0125788P.
PR   25-MAR-1999;   99US-0126264P.
PR   29-MAR-1999;   99US-0126785P.
PR   01-APR-1999;   99US-0127462P.
PR   06-APR-1999;   99US-0128234P.
PR   08-APR-1999;   99US-0128714P.
PR   16-APR-1999;   99US-0129845P.
PR   19-APR-1999;   99US-0130077P.
PR   21-APR-1999;   99US-0130449P.
PR   23-APR-1999;   99US-0130510P.
PR   23-APR-1999;   99US-0130891P.
PR   28-APR-1999;   99US-0131449P.
PR   30-APR-1999;   99US-0132048P.
PR   30-APR-1999;   99US-0132407P.
```

```
PR   04-MAY-1999;   99US-0132484P.
PR   05-MAY-1999;   99US-0132485P.
PR   06-MAY-1999;   99US-0132486P.
PR   06-MAY-1999;   99US-0132487P.
PR   07-MAY-1999;   99US-0132863P.
PR   11-MAY-1999;   99US-0134256P.
PR   14-MAY-1999;   99US-0134218P.
PR   14-MAY-1999;   99US-0134219P.
PR   14-MAY-1999;   99US-0134422P.
PR   14-MAY-1999;   99US-0134370P.
PR   16-MAY-1999;   99US-0134768P.
PR   19-MAY-1999;   99US-0134941P.
PR   20-MAY-1999;   99US-0135124P.
PR   21-MAY-1999;   99US-0135353P.
PR   24-MAY-1999;   99US-0135629P.
PR   25-MAY-1999;   99US-0136021P.
PR   27-MAY-1999;   99US-0136392P.
PR   28-MAY-1999;   99US-0136782P.
PR   01-JUN-1999;   99US-0137222P.
PR   03-JUN-1999;   99US-0137529P.
PR   04-JUN-1999;   99US-0137502P.
PR   07-JUN-1999;   99US-0137724P.
PR   08-JUN-1999;   99US-0138094P.
PR   10-JUN-1999;   99US-0138540P.
PR   10-JUN-1999;   99US-0138847P.
PR   14-JUN-1999;   99US-0139119P.
PR   16-JUN-1999;   99US-0139452P.
PR   16-JUN-1999;   99US-0139453P.
PR   17-JUN-1999;   99US-0139492P.
PR   18-JUN-1999;   99US-0139454P.
PR   18-JUN-1999;   99US-0139455P.
PR   18-JUN-1999;   99US-0139456P.
PR   18-JUN-1999;   99US-0139457P.
PR   18-JUN-1999;   99US-0139458P.
PR   18-JUN-1999;   99US-0139459P.
PR   18-JUN-1999;   99US-0139460P.
PR   18-JUN-1999;   99US-0139461P.
PR   18-JUN-1999;   99US-0139462P.
PR   18-JUN-1999;   99US-0139463P.
PR   18-JUN-1999;   99US-0139750P.
PR   18-JUN-1999;   99US-0139763P.
PR   21-JUN-1999;   99US-0139817P.
PR   22-JUN-1999;   99US-0139899P.
PR   23-JUN-1999;   99US-0140353P.
PR   23-JUN-1999;   99US-0140354P.
PR   24-JUN-1999;   99US-0140695P.
PR   28-JUN-1999;   99US-0140823P.
PR   29-JUN-1999;   99US-0140991P.
PR   30-JUN-1999;   99US-0141287P.
PR   01-JUL-1999;   99US-0141842P.
PR   01-JUL-1999;   99US-0142154P.
PR   02-JUL-1999;   99US-0142205P.
PR   06-JUL-1999;   99US-0142390P.
PR   08-JUL-1999;   99US-0142803P.
PR   09-JUL-1999;   99US-0142920P.
PR   12-JUL-1999;   99US-0142977P.
PR   13-JUL-1999;   99US-0143542P.
PR   14-JUL-1999;   99US-0143624P.
PR   15-JUL-1999;   99US-0144005P.
PR   16-JUL-1999;   99US-0144008P.
PR   16-JUL-1999;   99US-0144086P.
PR   19-JUL-1999;   99US-0144325P.
PR   19-JUL-1999;   99US-0144331P.
PR   19-JUL-1999;   99US-0144332P.
PR   19-JUL-1999;   99US-0144333P.
PR   19-JUL-1999;   99US-0144334P.
PR   19-JUL-1999;   99US-0144335P.
PR   20-JUL-1999;   99US-0144352P.
PR   20-JUL-1999;   99US-0144632P.
PR   20-JUL-1999;   99US-0144884P.
PR   21-JUL-1999;   99US-0144814P.
PR   21-JUL-1999;   99US-0145086P.
PR   21-JUL-1999;   99US-0145088P.
```

```
PR   05-MAY-1999;   99US-0132485P.;
PR   06-MAY-1999;   99US-0132486P.;
PR   06-MAY-1999;   99US-0132487P.;
PR   07-MAY-1999;   99US-0132863P.;
PR   11-MAY-1999;   99US-0134256P.;
PR   14-MAY-1999;   99US-0134218P.;
PR   14-MAY-1999;   99US-0134219P.;
PR   14-MAY-1999;   99US-0134221P.;
PR   14-MAY-1999;   99US-0134370P.;
PR   18-MAY-1999;   99US-0134768P.;
PR   19-MAY-1999;   99US-0134941P.;
PR   20-MAY-1999;   99US-0135124P.;
PR   21-MAY-1999;   99US-0135353P.;
PR   24-MAY-1999;   99US-0135629P.;
PR   25-MAY-1999;   99US-0136021P.;
PR   27-MAY-1999;   99US-0136392P.;
PR   28-MAY-1999;   99US-0136782P.;
PR   01-JUN-1999;   99US-0137222P.;
PR   03-JUN-1999;   99US-0137528P.;
PR   04-JUN-1999;   99US-0137502P.;
PR   07-JUN-1999;   99US-0137724P.;
PR   08-JUN-1999;   99US-0138094P.;
PR   10-JUN-1999;   99US-0138540P.;
PR   10-JUN-1999;   99US-0138847P.;
PR   14-JUN-1999;   99US-0139119P.;
PR   16-JUN-1999;   99US-0139452P.;
PR   16-JUN-1999;   99US-0139453P.;
PR   17-JUN-1999;   99US-0139492P.;
PR   18-JUN-1999;   99US-0139454P.;
PR   18-JUN-1999;   99US-0139455P.;
PR   18-JUN-1999;   99US-0139456P.;
PR   18-JUN-1999;   99US-0139457P.;
PR   18-JUN-1999;   99US-0139458P.;
PR   18-JUN-1999;   99US-0139459P.;
PR   18-JUN-1999;   99US-0139460P.;
PR   18-JUN-1999;   99US-0139461P.;
PR   18-JUN-1999;   99US-0139462P.;
PR   18-JUN-1999;   99US-0139463P.;
PR   18-JUN-1999;   99US-0139750P.;
PR   18-JUN-1999;   99US-0139763P.;
PR   21-JUN-1999;   99US-0139817P.;
PR   22-JUN-1999;   99US-0139899P.;
PR   23-JUN-1999;   99US-0140353P.;
PR   23-JUN-1999;   99US-0140354P.;
PR   24-JUN-1999;   99US-0140695P.;
PR   28-JUN-1999;   99US-0140823P.;
PR   29-JUN-1999;   99US-0140991P.;
PR   30-JUN-1999;   99US-0141287P.;
PR   01-JUL-1999;   99US-0141842P.;
PR   01-JUL-1999;   99US-0142154P.;
PR   02-JUL-1999;   99US-0142055P.;
PR   06-JUL-1999;   99US-0142390P.;
PR   08-JUL-1999;   99US-0142803P.;
PR   09-JUL-1999;   99US-0142920P.;
PR   12-JUL-1999;   99US-0142777P.;
PR   13-JUL-1999;   99US-0143542P.;
PR   14-JUL-1999;   99US-0143624P.;
PR   15-JUL-1999;   99US-0144005P.;
PR   16-JUL-1999;   99US-0144085P.;
PR   16-JUL-1999;   99US-0144086P.;
PR   19-JUL-1999;   99US-0144325P.;
PR   19-JUL-1999;   99US-0144331P.;
PR   19-JUL-1999;   99US-0144332P.;
PR   19-JUL-1999;   99US-0144333P.;
PR   19-JUL-1999;   99US-0144334P.;
PR   19-JUL-1999;   99US-0144335P.;
PR   20-JUL-1999;   99US-0144352P.;
PR   20-JUL-1999;   99US-0144632P.;
PR   21-JUL-1999;   99US-0144884P.;
PR   21-JUL-1999;   99US-0144814P.;
PR   21-JUL-1999;   99US-0145086P.;
PR   21-JUL-1999;   99US-0145088P.;
PR   22-JUL-1999;   99US-0145085P.;

PR   22-JUL-1999;   99US-0145087P.;
PR   22-JUL-1999;   99US-0145089P.;
PR   22-JUL-1999;   99US-0145192P.;
PR   23-JUL-1999;   99US-0145145P.;
PR   23-JUL-1999;   99US-0145218P.;
PR   23-JUL-1999;   99US-0145224P.;
PR   26-JUL-1999;   99US-0145276P.;
PR   27-JUL-1999;   99US-0145913P.;
PR   27-JUL-1999;   99US-0145918P.;
PR   27-JUL-1999;   99US-0145919P.;
PR   28-JUL-1999;   99US-0145951P.;
PR   02-AUG-1999;   99US-0146386P.;
PR   02-AUG-1999;   99US-0146388P.;
PR   03-AUG-1999;   99US-0146389P.;
PR   04-AUG-1999;   99US-0147038P.;
PR   04-AUG-1999;   99US-0147204P.;
PR   05-AUG-1999;   99US-0147302P.;
PR   05-AUG-1999;   99US-0147192P.;
PR   06-AUG-1999;   99US-0147260P.;
PR   06-AUG-1999;   99US-0147303P.;
PR   06-AUG-1999;   99US-0147416P.;
PR   09-AUG-1999;   99US-0147493P.;
PR   09-AUG-1999;   99US-0147935P.;
PR   10-AUG-1999;   99US-0148171P.;
PR   11-AUG-1999;   99US-0148319P.;
PR   12-AUG-1999;   99US-0148341P.;
PR   13-AUG-1999;   99US-0148565P.;
PR   13-AUG-1999;   99US-0148684P.;
PR   16-AUG-1999;   99US-0149368P.;
PR   17-AUG-1999;   99US-0149175P.;
PR   18-AUG-1999;   99US-0149426P.;
PR   20-AUG-1999;   99US-0149722P.;
PR   20-AUG-1999;   99US-0149723P.;
PR   20-AUG-1999;   99US-0149929P.;
PR   23-AUG-1999;   99US-0149902P.;
PR   25-AUG-1999;   99US-0149930P.;
PR   26-AUG-1999;   99US-0150566P.;
PR   27-AUG-1999;   99US-0150884P.;
PR   27-AUG-1999;   99US-0151065P.;
PR   30-AUG-1999;   99US-0151066P.;
PR   31-AUG-1999;   99US-0151080P.;
PR   01-SEP-1999;   99US-0151303P.;
PR   07-SEP-1999;   99US-0151438P.;
PR   10-SEP-1999;   99US-0151930P.;
PR   13-SEP-1999;   99US-0152363P.;
PR   15-SEP-1999;   99US-0153070P.;
PR   16-SEP-1999;   99US-0153758P.;
PR   20-SEP-1999;   99US-0154018P.;
PR   22-SEP-1999;   99US-0154039P.;
PR   23-SEP-1999;   99US-0154779P.;
PR   24-SEP-1999;   99US-0155139P.;
PR   28-SEP-1999;   99US-0155486P.;
PR   29-SEP-1999;   99US-0155659P.;
PR   04-OCT-1999;   99US-0156458P.;
PR   05-OCT-1999;   99US-0155696P.;
PR   06-OCT-1999;   99US-0157117P.;
PR   07-OCT-1999;   99US-0157753P.;
PR   08-OCT-1999;   99US-0157865P.;
PR   12-OCT-1999;   99US-0158029P.;
PR   13-OCT-1999;   99US-0158232P.;
PR   13-OCT-1999;   99US-0158369P.;
PR   14-OCT-1999;   99US-0159293P.;
PR   14-OCT-1999;   99US-0159294P.;
PR   14-OCT-1999;   99US-0159295P.;
PR   14-OCT-1999;   99US-0159329P.;
PR   18-OCT-1999;   99US-0159330P.;
PR   21-OCT-1999;   99US-0159331P.;
PR   21-OCT-1999;   99US-0159637P.;
PR   21-OCT-1999;   99US-0159638P.;
PR   21-OCT-1999;   99US-0159584P.;
PR   21-OCT-1999;   99US-0160741P.;
PR   21-OCT-1999;   99US-0160767P.;
PR   21-OCT-1999;   99US-0160768P.;
```

```
PR  21-OCT-1999;  99US-0160814P.
PR  21-OCT-1999;  99US-0160815P.
PR  22-OCT-1999;  99US-0160980P.
PR  22-OCT-1999;  99US-0160981P.
PR  22-OCT-1999;  99US-0160989P.
PR  25-OCT-1999;  99US-0161404P.
PR  25-OCT-1999;  99US-0161405P.
PR  25-OCT-1999;  99US-0161406P.
PR  26-OCT-1999;  99US-0161359P.
PR  26-OCT-1999;  99US-0161360P.
PR  26-OCT-1999;  99US-0161361P.
PR  28-OCT-1999;  99US-0161920P.
PR  28-OCT-1999;  99US-0161992P.
PR  28-OCT-1999;  99US-0161993P.
PR  29-OCT-1999;  99US-0162142P.


Query Match     14.6%;   Score 137.5;   DB 3;   Length 350;
Best Local Similarity   25.0%;   Pred. No. 3.6e-05;
Matches   45;  Conservative   26;  Mismatches   88;  Indels   21;  Gaps   6;

Qy    1 CPLQGSHALCTCCFQPMDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
          | |:|    ||||        |          :      | ||    :  |  |
Db  152 CRHHGAHLQCHLCQGMPFRANLQ------VPLHCKGCDRPFCGAYWS-SENVTQGVSGP 204

Qy   61 FCELN----LGDKCLDGV---LNNNSYESDILKNYLATRGLTWKQMLTESL-----VAL 107
          :|          :|:    ||| ::   :: :  :::     ||::
Db  205 VCVRETFRPISERTITRIPFITHEMNRHEQDITQRCIAHMEKTVPDVVAEWLRLFNNREI 264

Qy  108 QRGVFLLSDYR-VTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
          :: : :: :    |:|:|   :         ||       :|:      |
Db  265 DRSRMPLNHAEMITASTHVCNDCYDKLVGFLLYWFRITLPRNHLPADVAAREDCWYGYAC 324


RESULT 9
AAG37660
ID   AAG37660 standard; protein; 350 AA.
XX
AC   AAG37660;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 46344.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;  2000EP-00301439.
XX
PR   25-FEB-1999;  99US-0121825P.
PR   05-MAR-1999;  99US-0123180P.
PR   09-MAR-1999;  99US-0123548P.
PR   23-MAR-1999;  99US-0125788P.
PR   25-MAR-1999;  99US-0126264P.
PR   29-MAR-1999;  99US-0126785P.
PR   01-APR-1999;  99US-0127462P.
PR   06-APR-1999;  99US-0128234P.
PR   08-APR-1999;  99US-0128714P.
PR   16-APR-1999;  99US-0129845P.
PR   19-APR-1999;  99US-0130077P.
PR   21-APR-1999;  99US-0130449P.
PR   23-APR-1999;  99US-0130510P.
PR   23-APR-1999;  99US-0130891P.
PR   28-APR-1999;  99US-0131449P.
PR   30-APR-1999;  99US-0132048P.
PR   30-APR-1999;  99US-0132207P.
PR   04-MAY-1999;  99US-0132484P.
```

```
PR  22-JUL-1999;  99US-0145089P.
PR  22-JUL-1999;  99US-0145192P.
PR  23-JUL-1999;  99US-0145145P.
PR  23-JUL-1999;  99US-0145218P.
PR  23-JUL-1999;  99US-0145224P.
PR  26-JUL-1999;  99US-0145276P.
PR  26-JUL-1999;  99US-0145913P.
PR  27-JUL-1999;  99US-0145918P.
PR  27-JUL-1999;  99US-0145919P.
PR  28-JUL-1999;  99US-0145951P.
PR  02-AUG-1999;  99US-0146386P.
PR  02-AUG-1999;  99US-0146388P.
PR  02-AUG-1999;  99US-0146389P.
PR  03-AUG-1999;  99US-0147038P.
PR  03-AUG-1999;  99US-0147204P.
PR  04-AUG-1999;  99US-0147302P.
PR  05-AUG-1999;  99US-0147192P.
PR  05-AUG-1999;  99US-0147260P.
PR  06-AUG-1999;  99US-0147303P.
PR  06-AUG-1999;  99US-0147416P.
PR  09-AUG-1999;  99US-0147493P.
PR  09-AUG-1999;  99US-0147935P.
PR  10-AUG-1999;  99US-0148171P.
PR  11-AUG-1999;  99US-0148319P.
PR  12-AUG-1999;  99US-0148341P.
PR  13-AUG-1999;  99US-0148565P.
PR  13-AUG-1999;  99US-0148684P.
PR  16-AUG-1999;  99US-0149168P.
PR  17-AUG-1999;  99US-0149175P.
PR  18-AUG-1999;  99US-0149426P.
PR  20-AUG-1999;  99US-0149722P.
PR  20-AUG-1999;  99US-0149723P.
PR  20-AUG-1999;  99US-0149929P.
PR  23-AUG-1999;  99US-0149902P.
PR  23-AUG-1999;  99US-0149930P.
PR  25-AUG-1999;  99US-0150566P.
PR  26-AUG-1999;  99US-0150884P.
PR  27-AUG-1999;  99US-0151065P.
PR  27-AUG-1999;  99US-0151066P.
PR  30-AUG-1999;  99US-0151080P.
PR  31-AUG-1999;  99US-0151303P.
PR  31-AUG-1999;  99US-0151438P.
PR  01-SEP-1999;  99US-0151930P.
PR  07-SEP-1999;  99US-0152363P.
PR  10-SEP-1999;  99US-0153070P.
PR  13-SEP-1999;  99US-0153758P.
PR  15-SEP-1999;  99US-0154018P.
PR  16-SEP-1999;  99US-0154039P.
PR  20-SEP-1999;  99US-0154779P.
PR  22-SEP-1999;  99US-0155139P.
PR  23-SEP-1999;  99US-0155486P.
PR  24-SEP-1999;  99US-0155659P.
PR  28-SEP-1999;  99US-0156458P.
PR  29-SEP-1999;  99US-0156596P.
PR  04-OCT-1999;  99US-0157117P.
PR  05-OCT-1999;  99US-0157753P.
PR  06-OCT-1999;  99US-0157865P.
PR  07-OCT-1999;  99US-0158029P.
PR  08-OCT-1999;  99US-0158232P.
PR  12-OCT-1999;  99US-0158669P.
PR  13-OCT-1999;  99US-0159293P.
PR  13-OCT-1999;  99US-0159294P.
PR  13-OCT-1999;  99US-0159295P.
PR  14-OCT-1999;  99US-0159329P.
PR  14-OCT-1999;  99US-0159330P.
PR  14-OCT-1999;  99US-0159331P.
PR  14-OCT-1999;  99US-0159637P.
PR  14-OCT-1999;  99US-0159638P.
PR  18-OCT-1999;  99US-0159584P.
PR  21-OCT-1999;  99US-0160741P.
PR  21-OCT-1999;  99US-0160767P.
PR  21-OCT-1999;  99US-0160768P.
PR  21-OCT-1999;  99US-0160770P.
```

PI   Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR   WPI; 2003-395539/38.
DR   N-PSDB; ADA52592.
XX
PT   New polynucleotides encoding full-length polypeptides, e.g. secretory
PT   and/or membrane proteins, useful for developing medicines for diseases in
PT   which the gene is involved, or as target molecules for gene therapy.
XX
PS   Claim 14; SEQ ID NO 1799; 205pp; English.
XX
CC   The present invention relates to novel human secretory or membrane
CC   proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC   ADA54071). The coding sequences are useful in the gene therapy of
CC   diseases caused by abnormalities of the proteins, e.g. cancer,
CC   inflammatory diseases, osteoporosis or neurological disease.
XX
SQ   Sequence 128 AA;

Query Match   46.4%; Score 438; DB 6; Length 128;
Best Local Similarity   98.8%; Pred. No. 2.1e-35;
Matches   80; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy   86 KNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNI 145
          ::
Db   25 QNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNI 84

Qy   146 PASELPVAVTSRPDCYWGRNC 166
          ::::::::::::::::::::::
Db   85 PASELPVAVTSRPDCYWGRNC 105

RESULT 8
AAG20354
ID   AAG20354 standard; protein; 350 AA.
XX
AC   AAG20354;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 22507.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX

PR	25-FEB-1999;	99US-0121825P.	PR	06-MAY-1999;	99US-0132486P.
PR	05-MAR-1999;	99US-0123180P.	PR	06-MAY-1999;	99US-0132487P.
PR	09-MAR-1999;	99US-0123548P.	PR	07-MAY-1999;	99US-0132863P.
PR	23-MAR-1999;	99US-0125788P.	PR	11-MAY-1999;	99US-0134256P.
PR	25-MAR-1999;	99US-0126264P.	PR	14-MAY-1999;	99US-0134219P.
PR	29-MAR-1999;	99US-0126785P.	PR	14-MAY-1999;	99US-0134221P.
PR	01-APR-1999;	99US-0127462P.	PR	14-MAY-1999;	99US-0134370P.
PR	06-APR-1999;	99US-0128234P.	PR	18-MAY-1999;	99US-0134768P.
PR	08-APR-1999;	99US-0128714P.	PR	19-MAY-1999;	99US-0134941P.
PR	16-APR-1999;	99US-0129845P.	PR	20-MAY-1999;	99US-0135124P.
PR	19-APR-1999;	99US-0130077P.	PR	21-MAY-1999;	99US-0135353P.
PR	21-APR-1999;	99US-0130449P.	PR	24-MAY-1999;	99US-0135629P.
PR	23-APR-1999;	99US-0130510P.	PR	25-MAY-1999;	99US-0136021P.
PR	23-APR-1999;	99US-0130891P.	PR	27-MAY-1999;	99US-0136392P.
PR	28-APR-1999;	99US-0131449P.	PR	28-MAY-1999;	99US-0136782P.
PR	30-APR-1999;	99US-0132048P.	PR	01-JUN-1999;	99US-0137222P.
PR	30-APR-1999;	99US-0132407P.	PR	03-JUN-1999;	99US-0137528P.
PR	04-MAY-1999;	99US-0132484P.	PR	04-JUN-1999;	99US-0137502P.
PR	05-MAY-1999;	99US-0132485P.	PR	07-JUN-1999;	99US-0137724P.
PR			PR	08-JUN-1999;	99US-0138094P.
PR			PR	10-JUN-1999;	99US-0138540P.
PR			PR	10-JUN-1999;	99US-0138847P.
PR			PR	14-JUN-1999;	99US-0139119P.
PR			PR	16-JUN-1999;	99US-0139452P.
PR			PR	16-JUN-1999;	99US-0139453P.
PR			PR	17-JUN-1999;	99US-0139492P.
PR			PR	18-JUN-1999;	99US-0139454P.
PR			PR	18-JUN-1999;	99US-0139455P.
PR			PR	18-JUN-1999;	99US-0139456P.
PR			PR	18-JUN-1999;	99US-0139457P.
PR			PR	18-JUN-1999;	99US-0139458P.
PR			PR	18-JUN-1999;	99US-0139459P.
PR			PR	18-JUN-1999;	99US-0139460P.
PR			PR	18-JUN-1999;	99US-0139461P.
PR			PR	18-JUN-1999;	99US-0139462P.
PR			PR	18-JUN-1999;	99US-0139463P.
PR			PR	18-JUN-1999;	99US-0139750P.
PR			PR	18-JUN-1999;	99US-0139763P.
PR			PR	21-JUN-1999;	99US-0139817P.
PR			PR	22-JUN-1999;	99US-0139899P.
PR			PR	23-JUN-1999;	99US-0140353P.
PR			PR	23-JUN-1999;	99US-0140354P.
PR			PR	24-JUN-1999;	99US-0140695P.
PR			PR	28-JUN-1999;	99US-0140823P.
PR			PR	29-JUN-1999;	99US-0140991P.
PR			PR	30-JUN-1999;	99US-0141287P.
PR			PR	01-JUL-1999;	99US-0141842P.
PR			PR	01-JUL-1999;	99US-0142154P.
PR			PR	02-JUL-1999;	99US-0142055P.
PR			PR	06-JUL-1999;	99US-0142390P.
PR			PR	08-JUL-1999;	99US-0142803P.
PR			PR	09-JUL-1999;	99US-0142920P.
PR			PR	12-JUL-1999;	99US-0142977P.
PR			PR	13-JUL-1999;	99US-0143542P.
PR			PR	14-JUL-1999;	99US-0143624P.
PR			PR	15-JUL-1999;	99US-0144005P.
PR			PR	16-JUL-1999;	99US-0144085P.
PR			PR	16-JUL-1999;	99US-0144086P.
PR			PR	19-JUL-1999;	99US-0144325P.
PR			PR	19-JUL-1999;	99US-0144331P.
PR			PR	19-JUL-1999;	99US-0144332P.
PR			PR	19-JUL-1999;	99US-0144333P.
PR			PR	19-JUL-1999;	99US-0144334P.
PR			PR	19-JUL-1999;	99US-0144335P.
PR			PR	20-JUL-1999;	99US-0144352P.
PR			PR	20-JUL-1999;	99US-0144632P.
PR			PR	20-JUL-1999;	99US-0144884P.
PR			PR	21-JUL-1999;	99US-0144814P.
PR			PR	21-JUL-1999;	99US-0144331P.
PR			PR	21-JUL-1999;	99US-0145086P.
PR			PR	22-JUL-1999;	99US-0145088P.
PR			PR	22-JUL-1999;	99US-0145085P.
PR			PR	22-JUL-1999;	99US-0145087P.

PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 12128; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB2446 to AAB95893
CC represent human amino acid sequences; AAB2446 to AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 623 AA;

Query Match   99.6%; Score 939; DB 4; Length 623;
Best Local Similarity 99.4%; Pred. No. 3.5e-84;
Matches 165; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy   1 CPLQGSHALCTCCFQPMDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 435 CPLQGSHALCTCCFQPMDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 494

Qy   61 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 495 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 554

Qy 121 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 555 GDTVLCYCCGLRSFRELTYQYQQNIPASELPVAVTSRPDCYWGRNC 600

RESULT 6
ABB97233
ID ABB97233 standard; protein; 623 AA.
XX
AC ABB97233;
XX
DT 27-JUN-2002 (first entry)
XX
DE Novel human protein SEQ ID NO: 501.
XX
KW Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag.
XX
OS Homo sapiens.
XX
XX
PN WO200222660-A2.
XX
PD 21-MAR-2002.
XX
PR 10-SEP-2001; 2001WO-US026015.
PR 11-SEP-2000; 2000US-00659671.
XX

PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-292408/33.
DR N-PSDB; ABN32419.
XX
PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX
PS Example 2; SEQ ID NO 501; 509pp; English.
XX
CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention
XX
SQ Sequence 623 AA;

Query Match   99.6%; Score 939; DB 5; Length 623;
Best Local Similarity 99.4%; Pred. No. 3.5e-84;
Matches 165; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy   1 CPLQGSHALCTCCFQPMDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 435 CPLQGSHALCTCCFQPMDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 494

Qy   61 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 495 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 554

Qy 121 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 555 GDTVLCYCCGLRSFRELTYQYQQNIPASELPVAVTSRPDCYWGRNC 600

RESULT 7
ADA54231
ID ADA54231 standard; protein; 128 AA.
XX
AC ADA54231;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human protein, SEQ ID 1799.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;

CC prophase to metaphase during mitosis. It has ubiquitin-protein ligase
CC activity. The Chfr checkpoint was evident in primary human cells, but was
CC inactivated in 4 of 8 human cancer cell lines. In U2OS cells, a mutation
CC was identified that caused a Val to Met amino acid substitution in the
CC highly conserved C-terminal Cys-rich region of the Chfr protein. In the
CC absence of the Chfr checkpoint, cells subjected to mitotic stress
CC condensed their chromosomes despite failing to separate their
CC chromosomes. Chfr may monitor centrosome separation. Inactivation of the
CC chfr gene (see AAF30352) in human cancer is theorized to underlie the
CC increased sensitivity of cancer cells to antimitotic drugs. Polypeptides
CC comprising the present sequence, or sequences comprising at least amino
CC acids 31-103, 303-316 and/or 476-641 of this sequence, are claimed.
CC Claimed methods of determining the tumourigenic potential of a cell
CC comprise examining the cell for the presence of Chfr expression or for
CC the presence of Chfr-mediated ubiquitin-protein ligase activity (in both
CC cases, absence of expression indicating predisposition to tumourigenesis
CC upon exposure to mitotic stress). A diagnostic kit for detecting the
CC tumourigenic potential of cell cells comprise may comprise a ligand that
CC binds to Chfr, such as an antibody or its fragment. Inhibitors of Chfr
CC are identified by monitoring their effect on Chfr expression, and are
CC used to retard the growth of cancer cells
XX
SQ Sequence 664 AA;

Query Match   100.0%;   Score 943; DB 4; Length 664;
Best Local Similarity 100.0%; Pred. No. 1.5e-84;
Matches 166; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCRTGCYGCLAP 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 476 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCRTGCYGCLAP 535

Qy 61 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 536 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 595

Qy 121 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
        ||||||||||||||||||||||||||||||||||||||||||||||
Db 596 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 641

RESULT 4
AAO08972
ID AAO08972 standard; protein; 269 AA.
XX
AC AAO08972;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 22864.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO20164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
PR 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR N-PSDB; AAI88903.

XX
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
PS Claim 20; SEQ ID NO 22864; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 269 AA;

Query Match   99.7%;   Score 940; DB 4; Length 269;
Best Local Similarity 99.4%; Pred. No. 1e-84;
Matches 165; Conservative   1; Mismatches   1; Indels   0; Gaps   0;

Qy   1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCRTGCYGCLAP 60
        ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db 81 CPLQGSHALCTCCFQPMPDRRAERRQDPRVAPQQCAVCLQPFCHLYWGCRTGCYGCLAP 140

Qy 61 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 120
        |||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db 141 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLMALQRGVFLLSDYRVT 200

Qy 121 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
        ||||||||||||||||||||||||||||||||||||||||||||||
Db 201 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 246

RESULT 5
AAB93182
ID AAB93182 standard; protein; 623 AA.
XX
AC AAB93182;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12128.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999;   99JP-00248036.
PR 27-AUG-1999;   99JP-00300253.
PR 11-JAN-2000;   2000JP-00018776.
PR 02-MAY-2000;   2000JP-00183767.
PR 09-JUN-2000;   2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-

CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 652 AA;

Query Match   100.0%; Score 943; DB 4; Length 652;
Best Local Similarity   100.0%; Pred. No. 1.5e-84;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 464 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 523

Qy   61 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 524 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 583

Qy 121 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 584 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 629

RESULT 2
AAB83843
ID AAB83843 standard; protein; 664 AA.
XX
AC AAB83843;
XX
DT 22-AUG-2001 (first entry)
XX
DE Amino acid sequence of a human ring finger protein designated FHAR1.
XX
KW FHAR1; RING finger protein; cancer; vaccine.
XX
OS Homo sapiens.
XX
PN WO200142430-A1.
XX
PD 14-JUN-2001.
XX
PF 07-DEC-2000; 2000WO-US033094.
XX
PR 08-DEC-1999; 99US-00456876.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Zhou B, Zhu Y, Chaturvedi P, Hurle MR, Li X;
XX
DR WPI; 2001-381663/40.
DR N-PSDB; AAF89709.
XX
PT New FHAR1 polypeptide, a member of the RING finger protein family for
PT diagnosing and treating cancer, and for use in anti-cancer vaccines.
XX
PS Claim 1; Page 19; 28pp; English.
XX
CC The present sequence represents a FHAR1 polypeptide, which is a member of
CC the RING finger protein family. FHAR1 is useful in the treatment of
CC cancer, and as a vaccine for inducing an immunological response in a
CC mammal. FHAR1 polynucleotides may also be used as a diagnostic reagent
CC through detection of mutations in the associated gene, and for chromosome

CC localization studies, and tissue expression studies. FHAR1 antibodies are
CC useful to isolate and to identify clones expressing the polypeptides, or
CC to purify the polypeptides by affinity chromatography and to treat cancer
XX
SQ Sequence 664 AA;

Query Match   100.0%; Score 943; DB 4; Length 664;
Best Local Similarity   100.0%; Pred. No. 1.5e-84;
Matches 166; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 476 CPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAP 535

Qy   61 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 536 FCELNLGDKCLDGVLNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT 595

Qy 121 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 166
         ||||||||||||||||||||||||||||||||||||||||||||||
Db 596 GDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNC 641

RESULT 3
AAB20219
ID AAB20219 standard; protein; 664 AA.
XX
AC AAB20219;
XX
DT 14-MAY-2001 (first entry)
XX
DE Human Chfr (checkpoint with FHA and ring finger) protein.
XX
KW Checkpoint with forkhead associated domain and ring finger; Chfr; human;
KW mitosis; cell cycle; tumour; diagnosis; antitumour; drug screening;
KW ubiquitin-protein ligase.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Domain   31..103
FT   /label= Forkhead-associated_domain
FT Domain   303..346
FT   /label= Ring_finger-domain
FT Region   476..641
FT   /note= "cysteine-rich region"
FT Misc-difference 580
FT   /note= "Met in U2OS cells"
XX
PN WO200109150-A2.
XX
PD 08-FEB-2001.
XX
PF 14-JUN-2000; 2000WO-US016391.
XX
PR 29-JUL-1999; 99US-0146194P.
XX
PA (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI Halazonetis T, Scolnick D;
XX
DR WPI; 2001-182927/18.
DR N-PSDB; AAF30352.
XX
PT Novel nucleic acid sequence of mitotic checkpoint gene encoding a
PT checkpoint with forkhead-associated domain and ring finger protein, for
PT diagnosing tumorigenic cells and in screening for anticancer drugs.
XX
PS Claim 8(a); Fig 4A-C; 85pp; English.
XX
CC The present sequence is that of human mitotic checkpoint protein Chfr,
CC having a forkhead associated domain (FHA) and a ring finger domain. The
CC protein is required for regulation of the transition of cells from

```
              GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May  7, 2004, 14:36:57 ; Search time 34.0063 Seconds
                                       (without alignments)
                                       1379.240 Million cell updates/sec

Title:       US-10-048-046-2_COPY_476_641
Perfect score: 943
Sequence:    1 CPLQGSHALCTCCFQPMPDR.........ASELPVAVTSRPDCYWGRNC 166

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:    1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:   1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_29Jan04:*
          1:   geneseqp1980s:*
          2:   geneseqp1990s:*
          3:   geneseqp2000s:*
          4:   geneseqp2001s:*
          5:   geneseqp2002s:*
          6:   geneseqp2003as:*
          7:   geneseqp2003bs:*
          8:   geneseqp2004s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| | | % | | | | |
| Result | | Query | | | | |
No.	Score	Match	Length	DB	ID	Description
1	943	100.0	652	4	AAB93168	Aab93168 Human pro
2	943	100.0	664	4	AAB83843	Aab83843 Amino aci
3	943	100.0	664	4	AAB20219	Aab20219 Human Chf
4	940	99.7	269	4	AAO08972	Aao08972 Human pol
5	939	99.6	623	4	AAB93182	Aab93182 Human pro
6	939	99.6	623	4	ABB97233	Abb97233 Novel hum
7	438	46.4	128	6	ADA54231	Ada54231 Human pro
8	137.5	14.6	350	3	AAG20354	Aag20354 Arabidops
9	137.5	14.6	350	3	AAG37660	Aag37660 Arabidops
10	137.5	14.6	426	3	AAG20353	Aag20353 Arabidops
11	137.5	14.6	426	3	AAG37659	Aag37659 Arabidops
12	137.5	14.6	453	3	AAG20352	Aag20352 Arabidops
13	137.5	14.6	453	3	AAG37658	Aag37658 Arabidops
14	92.5	9.8	227	3	AAB34759	Aab34759 Human sec
15	85.5	9.1	244	4	AAU56425	Aau56425 Propionib
16	85.5	9.1	244	6	ABM52944	Abm52944 Propionib
17	85	9.0	229	4	AAM40223	Aam40223 Human pol
18	85	9.0	229	5	ABG34856	Abg34856 Human can
19	85	9.0	231	4	AAM42009	Aam42009 Human pol
20	85	9.0	504	4	AAE02617	Aae02617 Human hae
21	84	8.9	232	6	ABG73844	Abg73844 Mutant hu
22	82.5	8.7	1023	3	AAB11436	Aab11436 D. discoi
23	81	8.6	551	5	ABB04672	Abb04672 Desulfito
24	80	8.5	192	6	ABR55481	Abr55481 Amino aci
25	80	8.5	302	3	AAG47210	Aag47210 Arabidops
26	80	8.5	324	3	AAG47209	Aag47209 Arabidops
27	80	8.5	530	4	ABB71303	Abb71303 Drosophil
28	80	8.5	725	4	ABG18306	Abg18306 Novel hum
29	80	8.5	725	4	ABG28511	Abg28511 Novel hum
30	79.5	8.4	577	5	ABG96264	Abg96264 Human imm
31	79	8.4	317	7	ADE07954	Ade07954 Novel pro
32	79	8.4	1561	5	ABB97437	Abb97437 Novel hum
33	78.5	8.3	493	4	AAE02618	Aae02618 Human hae
34	77	8.2	232	6	ABG73838	Abg73838 Mutant hu
35	77	8.2	303	6	ABU92033	Abu92033 Human pro
36	76.5	8.1	418	4	AAB93353	Aab93353 Human pro
37	76.5	8.1	1700	4	ABB64608	Abb64608 Drosophil
38	76.5	8.1	2931	4	ABB68229	Abb68229 Drosophil
39	76	8.1	185	6	ADA48586	Ada48586 Rice prot
40	76	8.1	185	6	ADA47952	Ada47952 Rice prot
41	76	8.1	567	3	AAY54838	Aay54838 Human dia
42	76	8.1	567	4	AAE01864	Aae01864 Human dia
43	76	8.1	567	6	AAE36090	Aae36090 Human dia
44	76	8.1	567	6	AAE36089	Aae36089 Human dia
45	76	8.1	567	7	ADC77639	Adc77639 Human 994

ALIGNMENTS

```
RESULT 1
AAB93168
ID   AAB93168 standard; protein; 652 AA.
XX
AC   AAB93168;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:12100.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
PR   29-JUL-1999;   99JP-00248016.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000;   2000JP-00118776.
PR   02-MAY-2000;   2000JP-00183767.
PR   09-JUN-2000;   2000JP-00241899.
XX
PA   (HELI-) HELIX RBS INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesising polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification. Where a primer set comprises:
PT   (a) an oligo-dT primer and an oligonucleotide complementary to the
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 12100; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
```

RA   Sakai M., Aoki K., Ogura K., Kudoh Y., Kikuchi H., Zhang M.Q.,
RA   Yanagida M.;
RT   "A 38 kb segment containing the cdc2 gene from the left arm of fission
RT   yeast chromosome II: sequence analysis and characterization of the
RT   genomic DNA and cDNAs encoded on the segment.";
RL   Yeast 16:71-80(2000).
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AB004537; BAA21416.1; -.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE   304 AA;   34283 MW; C87F5F5A24217F23 CRC64;

Query Match   40.4%; Score 107.5; DB 3; Length 304;
Best Local Similarity   40.0%; Pred. No. 2.5e-06;
Matches   16; Conservative   7; Mismatches   16; Indels   1; Gaps   1;

Qy   2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCR 41
              ||:.::|| ||||||:|| .:|| : ||||||| ||
Db   186 CSLCMEFIH-CPAATECGHIFCWSCINGWTSKKSECPLCR 224

Search completed: May 7, 2004, 14:49:04
Job time : 8.36536 secs

```
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA   Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT   "Full-length messenger RNA sequences greatly improve genome
RT   annotation.";
RL   Genome Biol. 0:0-0(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA   Feldmann K.;
RT   "Full-Length cDNA from Arabidopsis thaliana.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AY087766; AAM65302.1; -.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE  426 AA;  47807 MW;  4897AF27781EB8833 CRC64;

Query Match             42.7%;  Score  113.5; DB 10;  Length 426;
Best Local Similarity   46.5%;  Pred. No. 5.3e-07;
Matches   20;  Conservative    4;  Mismatches  18;  Indels    1;  Gaps    1;

Qy      2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
            |||| :|:|::|:.||:||:|||||||||||||| :||||||
Db    365 CAICQEKMHTPILLR-CKHMFCEDCVSEWFERERTCPLCRALV 406


RESULT 13
Q8L610
ID   Q8L610      PRELIMINARY;      PRT;   426 AA.
AC   Q8L610;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   At5G01960.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AY099711; AAM20562.1; -.
DR   EMBL; BT000282; AAN15601.1; -.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE  426 AA;  47791 MW;  B24BF52lE9F663D1 CRC64;

Query Match             42.7%;  Score  113.5; DB 10;  Length 426;
Best Local Similarity   46.5%;  Pred. No. 5.3e-07;
Matches   20;  Conservative    4;  Mismatches  18;  Indels    1;  Gaps    1;

Qy      2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
            |||| :|:|::|:.||:||:|||||||||||||| :||||||
Db    365 CAICQEKMHTPILLR-CKHMFCEDCVSEWFERERTCPLCRALV 406


RESULT 14
O37928
ID   O37928      PRELIMINARY;      PRT;   498 AA.
AC   O37928;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Immediate early protein.
OS   Feline herpesvirus (Felid herpesvirus 1).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=10334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-27;
RA   Sussman M.D., Maes R.K.;
RT   "Nucleotide sequence and characterization of feline herpesvirus 1
RT   homologs for ICP0, glycoprotein gL and uracil DNA glycosylase.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AF022391; AAB80763.1; -.
DR   HSSP; P28990; 1CHC.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE  498 AA;  55012 MW;  6BC58DED3B9C58C8 CRC64;

Query Match             41.4%;  Score  110; DB 12;  Length 498;
Best Local Similarity   39.5%;  Pred. No. 1.8e-06;
Matches   17;  Conservative    8;  Mismatches  18;  Indels    0;  Gaps    0;

Qy      2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
            |  |||:  |:|:. ..:..::||:||:||||:|: :||:|:||
Db      8 CPICLDPMNDLTFTMPCLHKFCYSCLSRWVGLNNKCPLCKTSV 50


RESULT 15
O13628
ID   O13628      PRELIMINARY;      PRT;   304 AA.
AC   O13628;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   PAS4 protein.
GN   PI036.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972 h-;
RX   MEDLINE=20089027; PubMed=10620777;
RA   Machida M., Yamazaki S., Kunihiro S., Tanaka T., Kushida N., Jinno K.,
RA   Haikawa Y., Yamazaki J., Yamamoto S., Sekine M., Oguchi A., Nagai Y.,
```

DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 829 AA; 91849 MW; 5C5B74906E2B621D CRC64;

Query Match   44.4%; Score 118; DB 3; Length 829;
Best Local Similarity   41.1%; Pred. No. 2.4e-07;
Matches   23; Conservative   6; Mismatches 15; Indels 12; Gaps 2;

Qy   1 TCIICQDLL------HDCVSLQ-------PCMHTFCAACYSGWMERSSLCPTCRCPV 44
           |||||: | ||: | :||||||:|||:|||| ::|||
Db   340 TCIICREEMRPWDPHDPVRLERTRAKKLPCGHILHQGCLKSWLERQQVCPTCRRPV 395

RESULT 9
Q8BZT0
ID   Q8BZT0   PRELIMINARY;   PRT;   178 AA.
AC   Q8BZT0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RING finger protein homolog.
GN   2310035N15Rik.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cecum;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The RIKEN Genome Exploration Research Group Phase I & II Team;
RA   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK033609; BAC28388.1; -.
DR   MGD; MGI:1916837; 2310035N15Rik.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE 178 AA; 20842 MW; 792FB74D2F0938A1 CRC64;

Query Match   43.2%; Score 115; DB 11; Length 178;
Best Local Similarity   41.5%; Pred. No. 1.6e-07;
Matches   17; Conservative   9; Mismatches 15; Indels 0; Gaps 0;

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCR 41
           |||||:||| | |::: ::|:||:| | :: :||||
Db   16 TCRLCSGYLIDATTVTECLHTFCRSCLVKYLEENNTCPTCR 56

RESULT 10
Q8BTQ0
ID   Q8BTQ0   PRELIMINARY;   PRT;   241 AA.
AC   Q8BTQ0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RING finger protein homolog (Similar to ring finger protein 3).
GN   2310035N15Rik.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NOD; TISSUE=Thymus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The RIKEN Genome Exploration Research Group Phase I & II Team;
RA   "Analysis of the mouse transcriptome based on functional annotation of

RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N-3; TISSUE=Breast tumor;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK089091; BAC40745.1; -.
DR   EMBL; BC049266; AAH49266.1; -.
DR   MGD; MGI:1916837; 2310035N15Rik.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE 241 AA; 28046 MW; 1F4645C6E7A3DEA5 CRC64;

Query Match   43.2%; Score 115; DB 11; Length 241;
Best Local Similarity   41.5%; Pred. No. 2e-07;
Matches   17; Conservative   9; Mismatches 15; Indels 0; Gaps 0;

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCR 41
           |||||:||| | |::: ::|:||:| | :: :||||
Db   16 TCRLCSGYLIDATTVTECLHTFCRSCLVKYLEENNTCPTCR 56

RESULT 11
O15262
ID   O15262   PRELIMINARY;   PRT;   247 AA.
AC   O15262;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Ring finger protein.
GN   RNF3A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Abdul-Rauf M., Dyer M.J.;
RT   "Interactions of the BCL7A protein with novel ring finger proteins.";
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AJ001019; CAA04477.1; -.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 247 AA; 28748 MW; AFA425732644EF8A CRC64;

Query Match   43.2%; Score 115; DB 4; Length 247;
Best Local Similarity   41.5%; Pred. No. 2.1e-07;
Matches   17; Conservative   9; Mismatches 15; Indels 0; Gaps 0;

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCR 41
           |||||:||| | |::: ::|:||:| | :: :||||
Db   21 TCRLCSGYLIDATTVTECLHTFCRSCLVKYLEENNTCPTCR 61

RESULT 12
Q8LAK0
ID   Q8LAK0   PRELIMINARY;   PRT;   426 AA.
AC   Q8LAK0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20388685; PubMed=10935642;
RA   Scolnick D.M., Halazonetis T.D.;
RT   "Chfr defines a mitotic stress checkpoint that delays entry into
RT   metaphase.";
RL   Nature 406:430-435(2000).
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AF170724; AAF91084.1; -.
DR   PDB; 1LGP; 07-AUG-02.
DR   PDB; 1LGQ; 07-AUG-02.
DR   Genew; HGNC:20455; CHFR.
DR   GO; GO:0007093; P:mitotic checkpoint; TAS.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE   664 AA;  73386 MW;  141A1E7FEFAE36A2 CRC64;

Query Match             100.0%;  Score 266;  DB 4;  Length 664;
Best Local Similarity   100.0%;  Pred. No. 5.1e-27;
  Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
              ||||||||||||||||||||||||||||||||||||||||||||
Db        303 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 346


RESULT 6
Q81OL3
ID   Q81OL3         PRELIMINARY;       PRT;   664 AA.
AC   Q81OL3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RIKEN cDNA 5730484M20 gene.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Limb;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC049792; AAH49792.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 2.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE   664 AA;  73871 MW;  D651BE3B463DEBB6 CRC64;

Query Match             100.0%;  Score 266;  DB 11;  Length 664;
Best Local Similarity   100.0%;  Pred. No. 5.1e-27;
  Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
              ||||||||||||||||||||||||||||||||||||||||||||
Db        302 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 345
```

```
RESULT 7
Q9SX88
ID   Q9SX88         PRELIMINARY;       PRT;   473 AA.
AC   Q9SX88;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   F16N3.15.
GN   F16N3.15.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. columbia;
RA   Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA   Lee J., Liu A., Li J., Kremenetskaia I., Gonzalez A., Altafi H.,
RA   Araujo R., Brooks S., Buehler B., Chao Q., Conn L., Conway A.B.,
RA   Dunn P., Hansen N., Huizar L., Khan S., Kim C., Palm C., Rowley D.,
RA   Shinn P., Walker M., Davis R.W., Ecker J.R., Federspiel N.A.,
RA   Theologis A.;
RT   "Arabidopsis thaliana chromosome 1 BAC F16N3 sequence.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AC007519; AAD46042.1; -.
DR   PIR; C96516; C96516.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE   473 AA;  53874 MW;  D17217493A81C991 CRC64;

Query Match             50.8%;  Score 135;  DB 10;  Length 473;
Best Local Similarity   46.9%;  Pred. No. 8.3e-10;
  Matches   23;  Conservative    5;  Mismatches   15;  Indels    6;  Gaps    1;

Qy          2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSS------LCPTCRCPV 44
              ||||    ||| ||  |||  |:|:|| |:||  ||  |  :||| |||
Db        145 CSICLNIWHDVVTAAPCLHNFCNGCFSEWMRSBEKHGHVLCPQCRTTV 193


RESULT 8
Q8NJ06
ID   Q8NJ06         PRELIMINARY;       PRT;   829 AA.
AC   Q8NJ06;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Conserved hypothetical protein.
GN   5F3.020.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Aign V., Hoheisel J., Brandt P., Hartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Mannhaupt G.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AL807371; CAD37003.1; -.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
```

```
Qy      1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||||||||||||||||||||||||||||||||||||||||||
Db    262 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 305

RESULT 2
Q96SL3
ID   Q96SL3;             PRELIMINARY;       PRT;   652 AA.
AC   Q96SL3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ14781.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AK027687; BAB55297.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE   652 AA;  72030 MW;  572F2C56D1743D80 CRC64;

Query Match             100.0%;  Score 266;  DB 4;  Length 652;
Best Local Similarity   100.0%;  Pred. No. 5.1e-27;
Matches    44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||||||||||||||||||||||||||||||||||||||||||
Db    291 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 334

RESULT 3
Q96EP1
ID   Q96EP1;             PRELIMINARY;       PRT;   652 AA.
AC   Q96EP1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; BC012072; AAH12072.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE   652 AA;  72058 MW;  1A50773CEA679F07 CRC64;

Query Match             100.0%;  Score 266;  DB 4;  Length 652;
Best Local Similarity   100.0%;  Pred. No. 5.1e-27;
Matches    44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||||||||||||||||||||||||||||||||||||||||||
Db    291 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 334

RESULT 4
Q8BJZ9
ID   Q8BJZ9;             PRELIMINARY;       PRT;   663 AA.
AC   Q8BJZ9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cell cycle checkpoint protein CHFR.
GN   5730484M20RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Body;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK077629; BAC36912.1; -.
DR   MGD; MGI:2444898; 5730484M20Rik.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 2.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE   663 AA;  73855 MW;  42BD653DCB4ADF45 CRC64;

Query Match             100.0%;  Score 266;  DB 11;  Length 663;
Best Local Similarity   100.0%;  Pred. No. 5.1e-27;
Matches    44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||||||||||||||||||||||||||||||||||||||||||
Db    302 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 345

RESULT 5
Q9NRT4
ID   Q9NRT4;             PRELIMINARY;       PRT;   664 AA.
AC   Q9NRT4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cell cycle checkpoint protein CHFR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

OM protein - protein search, using sw model

```
              GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.
```

Run on: May 7, 2004, 14:40:47 ; Search time 6.36536 Seconds
                                     (without alignments)
                                     2180.991 Million cell updates/sec

Title: US-10-048-046-2_COPY_303_346
Perfect score: 266
Sequence: 1 TCIICQDLLHDCVSLQPCMH.........ACYSGWMERSSLCPTCRCPV 44

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SPTREMBL_25:*
          1: sp_archea:*
          2: sp_bacteria:*
          3: sp_fungi:*
          4: sp_human:*
          5: sp_invertebrate:*
          6: sp_mammal:*
          7: sp_mhc:*
          8: sp_organelle:*
          9: sp_phage:*
         10: sp_plant:*
         11: sp_rodent:*
         12: sp_virus:*
         13: sp_vertebrate:*
         14: sp_unclassified:*
         15: sp_rvirus:*
         16: sp_bacteriap:*
         17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	266	100.0	623	4	Q9NVD5	Q9nvd5 homo sapien
2	266	100.0	652	4	Q96SL3	Q96sl3 homo sapien
3	266	100.0	652	4	Q96EP1	Q96ep1 homo sapien
4	266	100.0	663	4	Q8BJZ9	Q8bjz9 mus musculu
5	266	100.0	664	4	Q9NRT4	Q9nrt4 homo sapien
6	266	50.8	473	10	Q8IOL3	Q8iol3 mus musculu
7	135	50.0	473	10	Q9SX88	Q9sx88 arabidopsis
8	118	44.4	829	3	Q8NJO6	Q8nj06 neurospora
9	115	43.2	178	11	Q8BZT0	Q8bzt0 mus musculu
10	115	43.2	241	11	Q8BTQ0	Q8btq0 mus musculu
11	115	43.2	247	4	O15262	O15262 homo sapien
12	113.5	42.7	426	10	Q8LAK0	Q8lak0 arabidopsis
13	113.5	42.7	426	10	Q8L610	Q8l610 arabidopsis
14	110	41.4	498	12	O37928	O37928 feline herp
15	107.5	40.4	304	3	O13628	O13628 schizosacch
16	107.5	40.4	306	3	Q9UUF0	Q9uuf0 schizosacch
17	107	40.2	551	3	Q08109	Q08109 saccharomyc
18	105.5	39.7	515	10	Q9SIU7	Q9siu7 arabidopsis
19	104	39.1	610	5	Q20798	Q20798 caenorhabdi
20	104	39.1	618	13	Q80318	Q80318 brachydanio
21	104	39.1	639	5	Q21641	Q21641 caenorhabdi
22	103.5	38.9	540	13	Q72X20	Q72x20 xenopus lae
23	103	38.7	298	4	Q8TEH6	Q8teh6 homo sapien
24	103	38.7	575	11	Q80T88	Q80t88 mus musculu
25	103	38.7	579	4	Q96JU5	Q96ju5 homo sapien
26	103	38.7	612	11	Q8CGB5	Q8cgb5 mus musculu
27	103	38.7	612	11	Q9DBY1	Q9dby1 mus musculu
28	103	38.7	616	4	Q8N3K3	Q8n3k3 homo sapien
29	103	38.7	616	4	Q96PK3	Q96pk3 homo sapien
30	103	38.7	617	4	Q8N6E8	Q8n6e8 homo sapien
31	103	38.7	617	4	Q86TM6	Q86tm6 homo sapien
32	102	38.3	328	5	Q9VJW5	Q9vjw5 drosophila
33	102	38.3	351	5	Q9NKF3	Q9nkf3 drosophila
34	101.5	38.2	161	10	Q9FN58	Q9fn58 arabidopsis
35	101.5	38.2	237	4	Q96A37	Q96a37 homo sapien
36	101.5	38.2	459	3	Q7ZA17	Q7za17 yarrowia li
37	101	38.0	496	11	Q9D6L2	Q9d6l2 mus musculu
38	101	38.0	587	5	Q96182	Q96182 plasmodium
39	100.5	37.8	132	10	Q9M0R7	Q9m0r7 arabidopsis
40	100.5	37.8	486	3	O60106	O60106 schizosacch
41	100.5	37.8	496	10	Q9SS55	Q9ss55 arabidopsis
42	100.5	37.8	652	4	Q86Y15	Q86y15 homo sapien
43	100.5	37.8	673	12	Q67469	Q67469 feldmannia
44	100	37.8	1208	4	O75162	O75162 homo sapien
45	100	37.6	168	12	Q8QLD7	Q8qld7 mamestra co

ALIGNMENTS

RESULT 1
Q9NVD5
ID   Q9NVD5   PRELIMINARY;   PRT;   623 AA.
AC   Q9NVD5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ10796.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AK001658; BAA91817.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE   623 AA; 69204 MW; 45186D33DAE52711 CRC64;

Query Match   100.0%; Score 266; DB 4; Length 623;
Best Local Similarity   100.0%; Pred. No. 4.9e-27;
Matches   44; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Matches    16;  Conservative    6;  Mismatches   21;  Indels   2;  Gaps   1;

Qy        2  CIICQDLLHDCVSLQ--PCMHTFCAACYSGWMERSSLCPTCRCPV 44
             :||-:|     ||:|| ||||||||||||||||   ||: ||||||
Db      126  CAVCTDIAPPLRCQSFPCLHPFCIPCMKTWIPLRNTCPLCNTPV 170


RESULT 15
TRM3_MOUSE
ID   TRM3_MOUSE     STANDARD;      PRT;    744 AA.
AC   Q9R1R2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Tripartite motif protein 3 (RING finger protein 22) (RING finger
DE   protein HAC1).
GN   TRIM3 OR RNF22 OR HAC1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Yanai K., Shimamoto Y., Hirota K., Fukamizu A.;
RT   "Cloning of a new co-activator with ring finger motif.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21231161; PubMed=11331580;
RA   Reymond A., Meroni G., Fantozzi A., Merla G., Cairo S., Luzi L.,
RA   Riganelli D., Zanaria E., Messali S., Cainarca S., Guffanti A.,
RA   Minucci S., Pelicci P.G., Ballabio A.;
RT   "The tripartite motif family identifies cell compartments.";
RL   EMBO J. 20:2140-2151(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retina;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBUNIT: Associates with myosin V and alpha-actinin-4 (By
CC       similarity).
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- SIMILARITY: Contains 1 B box-type zinc finger.
CC   -!- SIMILARITY: Contains 1 filamin repeat.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long  as its content is  in  no  way
CC   modified and this statement is not removed. Usage  by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AB030912; BAA83343.1; -.
DR   EMBL; AF220019; AAG53473.1; -.
DR   EMBL; AK019165; BAB15801.1; -.
DR   EMBL; BC034263; AAH34263.1; -.
DR   HSSP; P29590; 1BOR.
DR   MGD; MGI:1860040; Trim3.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   InterPro; IPR003649; Bbox_C.
DR   InterPro; IPR001298; Filamin.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001258; NHL.
DR   InterPro; IPR000315; Znf_Bbox.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00630; Filamin; 1.
DR   Pfam; PF01436; NHL; 6.
DR   Pfam; PF00643; zf-B_box; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00502; BBC; 1.
DR   SMART; SM00336; BBOX; 1.
DR   SMART; SM00557; IG_FLMN; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50194; FILAMIN_REPEAT; 1.
DR   PROSITE; PS50119; ZF_BBOX_1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Zinc-finger; Coiled coil.
FT   ZN_FING      22      63       RING-TYPE.
FT   ZN_FING     110     151       B BOX-TYPE.
FT   DOMAIN      153     224       COILED COIL (POTENTIAL).
FT   REPEAT      317     418       FILAMIN.
SQ   SEQUENCE   744 AA;   80774 MW;   D9AEF4FA264BA168 CRC64;
```

```
Query Match                   36.1%;   Score 96;  DB 1;  Length 744;
Best Local Similarity         41.9%;   Pred. No. 0.00078;
Matches   18;  Conservative    5;  Mismatches   16;  Indels   4;  Gaps   2;

Qy     2  CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSL---CPTCR 41
          |  || :||| | |||:||| ::||||||||||  :  :  |||
Db    22  CSICLD-RYRCPKVLPCLHTFCERCLQNYIPPQSLTLSCPVCR 63


Search completed: May  7, 2004, 14:46:33
Job time : 2.4868 secs
```

```
ID   RN11_MOUSE     STANDARD;      PRT;   154 AA.
AC   Q9QYK7; Q9EQI1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RING finger protein 11 (Sid1669) (Nedd4 WW domain-binding protein 2).
GN   RNF11 OR N4WBP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20135600; PubMed=10673045;
RA   Seki N., Hattori A., Hayashi A., Kozuma S., Sasaki M., Suzuki Y.,
RA   Sugano S., Muramatsu M., Saito T.;
RT   "Cloning and expression profile of mouse and human genes, Rnf11/RNF11,
RT   encoding a novel RING-H2 finger protein.";
RL   Biochim. Biophys. Acta 1489:421-427(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20498735; PubMed=11042109;
RA   Jolliffe C.N., Harvey K.F., Haines B.P., Parasivam G., Kumar S.;
RT   "Identification of multiple proteins expressed in murine embryos as
RT   binding partners for the WW domains of the ubiquitin-protein ligase
RT   Nedd4.";
RL   Biochem. J. 351:557-565(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBUNIT: Interacts with NEDD4.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB024427; BAA84682.1; -.
DR   EMBL; AF220206; AAG44245.1; ALT_INIT.
DR   EMBL; BC010299; AAH10299.1; -.
DR   MGD; MGI:1352759; Rnf11.
DR   GO; GO:0000151; C:ubiquitin ligase complex; IPI.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IPI.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
DR   PROSITE; PS00089; ZF_RING_2; 1.
KW   Zinc-finger.
FT   ZN_FING       99    140       RING-TYPE.
SQ   SEQUENCE   154 AA;  17458 MW;  EF192AB0C2D4BF87 CRC64;

     Query Match             36.5%;  Score 97;  DB 1;  Length 154;
     Best Local Similarity   40.0%;  Pred. No. 0.00015;
     Matches  18;  Conservative   7;  Mismatches  18;  Indels  2;  Gaps  2;

Qy        2  CIIC-QDLLH-DCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV   44
              |||:| |:| |::::||:|||:|||:|:|:|     ::||:||
Db       99  CVICMMDFVYGDPIRFLPCMHIYHLDCIDDWLMRSFTCPSCMEPV  143
```

RESULT 14

```
ID   ICP0_HSV2H     STANDARD;      PRT;   825 AA.
AC   P28284;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Trans-acting transcriptional protein ICP0 (VMW118 protein).
GN   RL2.
OS   Herpes simplex virus (type 2 / strain HG52).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10315;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92113549; PubMed=1662697;
RA   McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT   "Comparative sequence analysis of the long repeat regions and
RT   adjoining parts of the long unique regions in the genomes of herpes
RT   simplex viruses types 1 and 2.";
RL   J. Gen. Virol. 72:3057-3075(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Dolan A.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D10471; BAA23427.1; -.
DR   EMBL; Z86099; CAB06760.1; -.
DR   PIR; JQ1501; EDBEXD.
DR   HSSP; P28990; 1CHC.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS00089; ZF_RING_2; 1.
KW   Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
KW   DNA-binding.
FT   DOMAIN       120    123       POLY-GLY.
FT   ZN_FING      126    167       RING-TYPE.
FT   DOMAIN       266    271       POLY-GLY.
FT   DOMAIN       292    295       POLY-SER.
FT   DOMAIN       342    345       POLY-ALA.
FT   DOMAIN       386    389       POLY-SER.
FT   DOMAIN       395    400       POLY-GLY.
FT   DOMAIN       425    428       POLY-ALA.
FT   DOMAIN       590    627       POLY-SER.
SQ   SEQUENCE   825 AA;  81986 MW;  5CEB15B58553A274 CRC64;

     Query Match             36.5%;  Score 97;  DB 1;  Length 825;
     Best Local Similarity   35.6%;  Pred. No. 0.00066;
```

```
RN   [3]
RP   STRUCTURE BY NMR OF 1-63.
RX   MEDLINE=94172642; PubMed=8126734;
RA   Barlow P.N., Luisi B., Milner A., Elliott M., Everett R.D.;
RT   "Structure of the C3HC4 domain by 1H-nuclear magnetic resonance
RT   spectroscopy. A new structural class of zinc-finger.";
RL   J. Mol. Biol. 237:201-211(1994).
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.
```
```
DR   EMBL; M86664; AAB02498.1; -.
DR   PIR; I36801; WZBEF5.
DR   PDB; 1CHC; 30-APR-94.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
KW   DNA-binding; 3D-structure.
FT   ZN_FING       8     47       RING-TYPE.
FT   METAL         8      8       ZINC 1.
FT   METAL        11     11       ZINC 1.
FT   METAL        24     24       ZINC 2.
FT   METAL        26     26       ZINC 2.
FT   METAL        29     29       ZINC 1.
FT   METAL        32     32       ZINC 1.
FT   METAL        43     43       ZINC 2.
FT   METAL        46     46       ZINC 2.
FT   DOMAIN      210    217       POLY-SER.
FT   STRAND       19     21
FT   TURN         22     25
FT   STRAND       26     28
FT   STRAND       30     31
FT   TURN         32     39
FT   HELIX        43     46
FT   TURN         53     54
FT   STRAND
SQ   SEQUENCE   532 AA;  58629 MW;  B4C87E16FA26FDFA CRC64;

Query Match             36.8%;  Score 98;  DB 1;  Length 532;
Best Local Similarity   34.9%;  Pred. No. 0.00034;
Matches   15;  Conservative   8;  Mismatches   20;  Indels   0;  Gaps   0;

Qy      2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
            :||.| |.|:||    |   :.|  |.:|| |::  ||  ||
Db      8 CPICLEDPSNYSMALPCLHAFCYVCITRWIRQNPTCPLCKVPV 50

RESULT 12
RN11_HUMAN
ID   RN11_HUMAN     STANDARD;      PRT;   154 AA.
AC   Q9Y3C5;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RING finger protein.11 (Sid1669) (CGI-123).
GN   RNF11.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20135600; PubMed=10673045;
RA   Seki N., Hattori A., Hayashi A., Kozuma S., Sasaki M., Suzuki Y.,
```

```
RA   Sugano S., Muramatsu M., Saito T.;
RT   "Cloning and expression profile of mouse and human genes, Rnf11/RNF11,
RT   encoding a novel RING-H2 finger protein.";
RL   Biochim. Biophys. Acta 1489:421-427(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20272150; PubMed=10810093;
RA   Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT   "Identification of novel human genes evolutionarily conserved in
RT   Caenorhabditis elegans by comparative proteomics.";
RL   Genome Res. 10:703-713(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Skin;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBUNIT: Interacts with NEDD4 (By similarity).
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
```
```
DR   EMBL; AB024703; BAA84683.1; -.
DR   EMBL; AF151881; AAD34118.1; -.
DR   EMBL; BC020964; AAH20964.1; -.
DR   EMBL; BC047654; AAH47654.1; -.
DR   Genew; HGNC:10056; RNF11.
DR   GO; GO:0003677; F:DNA binding; TAS.
DR   GO; GO:0008270; F:zinc ion binding; TAS.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Zinc-finger.
FT   ZN_FING      99    140       RING-TYPE.
SQ   SEQUENCE   154 AA;  17444 MW;  C368E3B148FC1D0D CRC64;

Query Match             36.5%;  Score 97;  DB 1;  Length 154;
Best Local Similarity   40.0%;  Pred. No. 0.0015;
Matches   18;  Conservative   7;  Mismatches   18;  Indels   2;  Gaps   2;

Qy      2 CIIC-QDLLH-DCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
            :||| | :|:  : ||: |||:|.|| |||.||:.||
Db     99 CVICMMDFVVGDPIRFLPCMHIYHLDCIDDWLMRSFTCPSCMEPV 143

RESULT 13
RN11_MOUSE
```

```
FT              /FTId=VSP_005759.
SQ   SEQUENCE   744 AA;  80836 MW;  3338151B7368B37E CRC64;

Query Match.
Best Local Similarity   37.2%;   Score 99;  DB 1;  Length 744;
                        41.9%;   Pred. No. 0.00036;
Matches  18;  Conservative   5;  Mismatches  16;  Indels  4;  Gaps  2;

Qy       2  CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSL---CPTCR 41
            | ||||| | |.||:|||      :       ||||
Db      22  CSICLD-RYQCPKVLPCLHTFCERCLQNYIPAQSLTLUSCPVCR 63

RESULT 10
CBLC_HUMAN
ID   CBLC_HUMAN     STANDARD;       PRT;    474 AA.
AC   Q9ULV8; Q9Y5Z2; Q9Y5Z3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Signal transduction protein CBL-C (SH3-binding protein CBL-C) (CBL-3).
GN   CBLC OR CBL3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM LONG).
RX   MEDLINE=20035821; PubMed=10571044;
RA   Kim H., Tezuka T., Suzuki Y., Sugano S., Hirai M., Yamamoto T.;
RT   "Molecular cloning and characterization of a novel cbl-family gene,
RT   cbl-c.";
RL   Gene 239:145-154(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC   TISSUE=Pancreatic adenocarcinoma;
RX   MEDLINE=99289203; PubMed=10362357;
RA   Keane M.M., Ettenberg S.A., Nau M.M., Banerjee P., Cuello M.,
RA   Penninger J., Lipkowitz S.;
RT   "cbl-3: a new mammalian cbl family protein.";
RL   Oncogene 18:3365-3375(1999).
CC   -!- FUNCTION: Regulator of EGFR mediated signal transduction.
CC   -!- SUBUNIT: Interacts with a restricted range of SH3 domain proteins.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=Q9ULV8-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=Q9ULV8-2; Sequence=VSP_005732;
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- DOMAIN: The N-terminus is composed of the phosphotyrosine binding
CC       (PTB) domain, a short linker region and the RING-type zinc finger.
CC       The PTB domain, which is also called TKB (tyrosine kinase binding)
CC       domain, is composed of three different subdomains: a four-helix
CC       bundle (4H), a calcium-binding EF hand and a divergent SH2 domain.
CC   -!- DOMAIN: The RING-type zinc finger domain mediates binding to an E2
CC       ubiquitin-conjugating enzyme (By similarity).
CC   -!- PTM: Phosphorylated on tyrosines by EGFR.
CC   -!- MISCELLANEOUS: This protein has one functional calcium-binding
CC       site (By similarity).
CC   -!- SIMILARITY: Contains 2 EF-hand-like calcium-binding domains.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- SIMILARITY: Contains 1 SH2 domain.
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/CBLcID194.html".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   --------------------------------------------------------------------------
DR   EMBL; AB028645; BAA86298.1; -.
DR   EMBL; AF117646; AAD34341.1; -.
DR   EMBL; AF117647; AAD34342.1; -.
DR   HSSP; P22681; 1B47.
DR   Genew; HGNC:15961; CBLC.
DR   GO; GO:0005515; F:protein binding; TAS.
DR   GO; GO:0008270; F:zinc ion binding; TAS.
DR   GO; GO:0007175; P:negative regulation of EGF receptor activity; TAS.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; TAS.
DR   InterPro; IPR003153; Cbl_N.
DR   InterPro; IPR000980; SH2.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF02761; Cbl_N2; 1.
DR   Pfam; PF02762; Cbl_N3; 1.
DR   Pfam; PF02262; Cbl_N; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   SMART; SM00252; SH2; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Nuclear protein; Zinc-finger; Phosphorylation; Alternative splicing;
KW   SH2 domain; Calcium-binding.
FT   DOMAIN        8    327       PTB.
FT   DOMAIN        8    147       4H.
FT   DOMAIN      182    190       EF-HAND LIKE 1.
FT   CA_BIND     199    210       EF-HAND LIKE 2.
FT   DOMAIN      237    311       SH2.
FT   DOMAIN      312    350       LINKER.
FT   ZN_FING     351    390       RING-TYPE.
FT   DOMAIN       89     92       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   BINDING     264    264       PHOSPHOTYROSINE (BY SIMILARITY).
FT   VARSPLIC    261    306       Missing (in isoform Short).
FT                                /FTId=VSP_005732.
FT   CONFLICT    234    234       T -> N (IN REF. 2).
SQ   SEQUENCE   474 AA;  52468 MW;  91013DDF12828242 CRC64;

Query Match                   36.8%;   Score 98;  DB 1;  Length 474;
Best Local Similarity         38.6%;   Pred. No. 0.00031;
Matches  17;  Conservative   6;  Mismatches  19;  Indels  2;  Gaps  2;

Qy       2  CIICQDLLHDCVSLQPCMHTFCAACYSGWMER-SSLCPTCRCPV 44
            || | :||||         :|:.|       |  |:| |||||
Db     351  CKICAESNKD-VKIEPCGHLLCSCCLAAWQHSDSQTCPFCRCBI 393

RESULT 11
ICP0_HSVEB
ID   ICP0_HSVEB     STANDARD;       PRT;    532 AA.
AC   P28990;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Trans-acting transcriptional protein ICP0.
GN   63.
OS   Equine herpesvirus type 1 (strain AB4p) (EHV-1).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=31520;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92295566; PubMed=1318606;
RA   Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT   "The DNA sequence of equine herpesvirus-1.";
RL   Virology 189:304-316(1992).
RN   [2]
RP   STRUCTURE BY NMR OF 1-63.
RX   MEDLINE=94087718; PubMed=8263911;
RA   Everett R.D., Barlow P.N., Milner A., Luisi B., Orr A., Hope G.,
RA   Lyon D.;
RT   "A novel arrangement of zinc-binding residues and secondary structure
RT   in the C3HC4 motif of an alpha herpes virus protein family.";
RL   J. Mol. Biol. 234:1038-1047(1993).
```

```
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.B.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Functions as an ubiquitin ligase protein in vitro.
CC   -!- SUBUNIT: Homodimer and heterodimer. Can heterodimerize with DTX1,
CC       enhancing its ubiquitin ligase activity in vitro. Interacts with
CC       BAL. Found in a complex with MYO6.
CC   -!- SIMILARITY: Belongs to the Deltex family.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   ----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------
DR   EMBL; AF484416; AAL90859.1; -.
DR   EMBL; BC042191; AAH42191.1; -.
DR   InterPro; IPR002350; kazal.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
FT   ZN_FING       561    600       RING-TYPE.
SQ   SEQUENCE   740 AA;  83554 MW;  C413BB744CEE6223 CRC64;

Query Match           37.2%;  Score 99;  DB 1;  Length 740;
Best Local Similarity 40.0%;  Pred. No. 0.00035;
Matches   16;  Conservative    6;  Mismatches   18;  Indels    0;  Gaps    0;

Qy     2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCR 41
          :|||: ||:  ||  : |:|   || |||   : ||||
Db   561 CVICMDTISNKKVLPCKKHEFCAPCINKAMSYKPICPTCQ 600

RESULT 9
TRM3_HUMAN
ID   TRM3_HUMAN     STANDARD;      PRT;    744 AA.
AC   O75382; Q9C038; Q9C039;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Tripartite motif protein 3 (RING finger protein 22) (Brain-expressed
DE   ring finger protein).
GN   TRM3 OR RNF22 OR BERP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC   TISSUE=Brain;
```

```
RX   MEDLINE=21100911; PubMed=11170753;
RA   El-Husseini A.E.-D., Fretier P., Vincent S.R.;
RT   "Cloning and characterization of a gene (RNF22) encoding a novel brain
RT   expressed ring finger protein (BERP) that maps to human chromosome
RT   11p15.5.";
RL   Genomics 71:363-367(2001).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS ALPHA; BETA AND GAMMA).
RX   MEDLINE=21231161; PubMed=11331580;
RA   Reymond A., Meroni G., Fantozzi A., Merla G., Cairo S., Luzi L.,
RA   Riganelli D., Zanaria E., Messali S., Cainarca S., Guffanti A.,
RA   Minucci S., Pelicci P.G., Ballabio A.;
RT   "The tripartite motif family identifies cell compartments.";
RL   EMBO J. 20:2140-2151(2001).
CC   -!- SUBUNIT: Associates with myosin V and alpha-actinin-4 (By
CC       similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=Alpha;
CC         IsoId=O75382-1; Sequence=Displayed;
CC       Name=Beta;
CC         IsoId=O75382-2; Sequence=VSP_005758;
CC       Name=Gamma;
CC         IsoId=O75382-3; Sequence=VSP_005759;
CC   -!- TISSUE SPECIFICITY: Expressed in brain, heart, uterus and testis.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- SIMILARITY: Contains 1 B box-type zinc finger.
CC   -!- SIMILARITY: Contains 1 filamin repeat.
CC   ----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------
DR   EMBL; AF045239; AAC24809.1; -.
DR   EMBL; AF220020; AAG53474.1; -.
DR   EMBL; AF220021; AAG53475.1; -.
DR   EMBL; AF220022; AAG53476.1; -.
DR   HSSP; P29590; 1BOR.
DR   Genew; HGNC:10064; TRIM3.
DR   MIM; 605493; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0008022; F:protein C-terminus binding; TAS.
DR   GO; GO:0007399; P:neurogenesis; TAS.
DR   InterPro; IPR003649; Bbox_C.
DR   InterPro; IPR001298; Filamin.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR001258; NHL.
DR   InterPro; IPR000315; Znf_Bbox.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00630; Filamin; 1.
DR   Pfam; PF01436; NHL; 6.
DR   Pfam; PF00643; zf-B_box; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00502; BBC; 1.
DR   SMART; SM00336; BBOX; 1.
DR   SMART; SM00557; IG_FLMN; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50194; FILAMIN_REPEAT; 1.
DR   PROSITE; PS50119; ZF_BBOX; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Zinc-finger; Coiled coil; Alternative splicing.
FT   ZN_FING        22     63       RING-TYPE.
FT   ZN_FING       110    151       B BOX-TYPE.
FT   DOMAIN        153    224       COILED COIL (POTENTIAL).
FT   REPEAT        317    418       FILAMIN.
FT   VARSPLIC      146    156       Missing (in isoform Beta).
FT                                  /FTId=VSP_005758.
FT   VARSPLIC      224    302       Missing (in isoform Gamma).
```

RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SIMILARITY: Contains 1 FHA domain.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; BC021778; AAH21778.1; -.
DR   MGD; MGI:1929069; Rnf8.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
FT   Pfam; PF00498; FHA; 1.
FT   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Zinc-finger.
FT   DOMAIN   38   92   FHA.
FT   DOMAIN   279   345   GLN-RICH.
FT   ZN_FING   406   444   RING-TYPE.
SQ   SEQUENCE   488 AA; 55516 MW; 428242204EBC44A1 CRC64;

Query Match   40.8%; Score 108.5; DB 1; Length 488;
Best Local Similarity   45.0%; Pred. No. 2e-05;
Matches   18; Conservative   7; Mismatches   14; Indels   1; Gaps   1;

Qy   2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCR 41
             ||| | ||:|:|| ||:||:|: | :|:||
Db   406 CIICSEYFIFAVTLN-CAHSFCSFCINEWMRKVECPICR 444

RESULT 7
YUES_CABEL
ID   YUES_CABEL   STANDARD;   PRT;   564 AA.
AC   P90859; P90852;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical RING finger protein F26E4.11 in chromosome I.
GN   F26E4.11
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]

RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Lightning J., Baynes C.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; Z81070; CAB03009.1; -.
DR   EMBL; Z81075; CAB03009.1; JOINED.
DR   EMBL; Z81075; CAB03049.1; -.
DR   EMBL; Z81070; CAB03049.1; JOINED.
DR   PIR; T21423; T21423.
DR   WormPep; F26E4.11; CE09695.
DR   InterPro; IPR003892; CUB.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF02845; CUB; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM05546; CUB; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING 1; FALSE_NEG.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Zinc-finger.
FT   ZN_FING   335   373   RING-TYPE.
SQ   SEQUENCE   564 AA; 64709 MW; 7C79CC238207E49B CRC64;

Query Match   38.9%; Score 103.5; DB 1; Length 564;
Best Local Similarity   41.9%; Pred. No. 8.5e-05;
Matches   18; Conservative   6; Mismatches   18; Indels   1; Gaps   1;

Qy   2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
             |::|:|| |: | || :| |:|: |:|:|| :|:
Db   335 CVVCWELLGTSRRL-PCSHQFHDWCLMWWLAQDSSCPTCRCTI 376

RESULT 8
BBAP_HUMAN
ID   BBAP_HUMAN   STANDARD;   PRT;   740 AA.
AC   Q8TDB6;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   B-lymphoma- and BAL-associated protein (Rhysin 2) (Rhysin2).
GN   BBAP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., IN VITRO UBIQUITIN LIGASE ACTIVITY, AND
RP   INTERACTION WITH BAL AND DTX1.
RX   MEDLINE=22679154; PubMed=12670957;
RA   Takeyama K., Aguiar R.C.T., Gu L., He C., Freeman G.J., Kutok J.L.,
RA   Aster J.C., Shipp M.A.;
RT   "The BAL-binding protein BBAP and related Deltex family members
RT   exhibit ubiquitin-protein isopeptide ligase activity.";
RL   J. Biol. Chem. 278:21930-21937(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Roberts R.C., Kendrick-Jones J., Jensen O.N.;
RT   "Rhysin2 is a novel protein identified by mass spectrometry found in a
RT   myosin VI-containing complex isolated by immunoprecipitation.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymph;
RX   MEDLINE=22388257; PubMed=12477932;

```
CC       -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U23511; AAC46798.2; -.
DR   WormPep; C3ZD5.10; CE29688.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; zinc-finger.
FT   ZN_FING      41     82       RING-TYPE.
FT   DOMAIN      451    499       ASP-RICH.
SQ   SEQUENCE   610 AA;  68276 MW;  60166BFD58AB806F CRC64;

Query Match             41.4%;  Score 110;  DB 1;  Length 610;
Best Local Similarity   40.0%;  Pred. No. 1.7e-05;
Matches   18;  Conservative    7;  Mismatches   18;  Indels    2;  Gaps     1;

Qy          2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSS--LCPTCRCPV 44
              |:|::|| |||||| | ||||  | |||:| || |:|
Db         41 CSVCKNEIIDTTSLSDCCHEFCYDCIVGWLTKGSGPFCPMCKTPV 85


RESULT 5
RNF8_HUMAN
ID   RNF8_HUMAN     STANDARD;      PRT;   485 AA.
AC   O76064;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RING finger protein 8.
GN   RNF8 OR KIAA0646.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Seki N., Yamauchi M., Saito T.;
RT   "Isolation and chromosomal assignment of the gene for a novel zinc
RT   finger protein.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC       -!- SIMILARITY: Contains 1 FHA domain.
CC       -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; AB012770; BAA33557.1; -.
DR   EMBL; AB014546; BAA31621.1; -.
DR   EMBL; AL096712; CAB75689.1; -.
DR   EMBL; BC007517; AAH07517.1; -.
DR   Genew; HGNC:10071; RNF8.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Zinc-finger.
FT   DOMAIN       38     92       FHA.
FT   DOMAIN      276    345       GLN-RICH.
FT   ZN_FING     403    441       RING-TYPE.
SQ   SEQUENCE   485 AA;  55517 MW;  5465OB2FFC9948B1 CRC64;

Query Match             40.8%;  Score 108.5;  DB 1;  Length 485;
Best Local Similarity   45.0%;  Pred. No. 2e-05;
Matches   18;  Conservative    7;  Mismatches   14;  Indels    1;  Gaps     1;

Qy          2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCR 41
              |||:|| ||||||:|||  | ||||||  ||:|| ||||
Db        403 CIICSEYFIEAVTLN-CAHSFCSYCINEWMKRKIECPICR 441


RESULT 6
RNF8_MOUSE
ID   RNF8_MOUSE     STANDARD;      PRT;   488 AA.
AC   Q8VC56;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RING finger protein 8.
GN   RNF8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
```

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
         ||||||..||...|.|.||
Db 18 TCICMSTVSDLGKTMPCLHDFCFVCIRAWTSTSVQCPLCRCPV 61

RESULT 2
ICP0_HSVBJ
ID ICP0_HSVBJ   STANDARD;   PRT;   676 AA.
AC P29128;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trans-acting transcriptional protein ICP0 (P135 protein) (IER
DE 2.9/ER2.6).
GN BICP0.
OS Bovine herpesvirus type 1 (strain Jura).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31518;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219360; PubMed=1313901;
RA Wirth U.V., Fraefel C., Vogt B., Vlcek C., Paces V., Schwyzer M.;
RT "Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1
RT are 3' coterminal and encode a putative zinc finger transactivator
RT protein.";
RL J. Virol. 66:2763-2772(1992).
CC -!- PTM: THE STRONGLY ACIDIC REGION MIGHT SERVE AS A TRANSCRIPTIONAL
CC   ACTIVATION DOMAIN, POSSIBLY REGULATED THROUGH PHOSPHORYLATION BY
CC   CASEIN KINASE II.
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M84465; AAA46062.1; -.
DR EMBL; AJ004801; CAA06138.1; -.
DR PIR; B38209; EDE23.
DR HSSP; P28999; 1CHC.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
KW DNA-binding; Early protein; Repressor; Phosphorylation.
FT ZN_FING   13   52   RING-TYPE.
FT DOMAIN   284   331   ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE   676 AA;   67879 MW; 11B06BA4E5C4EB71 CRC64;

Query Match 44.7%; Score 119; DB 1; Length 676;
Best Local Similarity 40.9%; Pred. No. 1.7e-06;
Matches 18; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
         :||:|||..||...|.|.||....|...|
Db 12 SCCICLDAITGAARALPCLHAFCLACIRRWLEGRPTCPLCKAPV 55

RESULT 3
ICP0_HSVBK
ID ICP0_HSVBK   STANDARD;   PRT;   676 AA.
AC P29836;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trans-acting transcriptional protein ICP0 (P135 protein) (IER

DE 2.9/ER2.6).
GN BICP0.
OS Bovine herpesvirus type 1 (strain K22).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31519;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219360; PubMed=1313901;
RA Wirth U.V., Fraefel C., Vogt B., Vlcek C., Paces V., Schwyzer M.;
RT "Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1
RT are 3' coterminal and encode a putative zinc finger transactivator
RT protein.";
RL J. Virol. 66:2763-2772(1992).
CC -!- PTM: THE STRONGLY ACIDIC REGION MIGHT SERVE AS A TRANSCRIPTIONAL
CC   ACTIVATION DOMAIN, POSSIBLY REGULATED THROUGH PHOSPHORYLATION BY
CC   CASEIN KINASE II.
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M84464; AAA46061.1; -.
DR PIR; A38209; EDE22.
DR HSSP; P28999; 1CHC.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
KW DNA-binding; Early protein; Repressor; Phosphorylation.
FT ZN_FING   13   52   RING-TYPE.
FT DOMAIN   284   331   ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE   676 AA;   67701 MW; 9BB0683C9BFCA65D CRC64;

Query Match 44.7%; Score 119; DB 1; Length 676;
Best Local Similarity 40.9%; Pred. No. 1.7e-06;
Matches 18; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
         :||:|||..||...|.|.||....|...|
Db 12 SCCICLDAITGAARALPCLHAFCLACIRRWLEGRPTCPLCKAPV 55

RESULT 4
YODA_CAEEL
ID YODA_CAEEL   STANDARD;   PRT;   610 AA.
AC Q09268;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical RING finger protein C32D5.10 in chromosome II.
GN C32D5.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Miller N.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.

```
              GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.
```

OM protein - protein search, using sw model

Run on: May 7, 2004, 14:37:32 ; Search time 1.4868 Seconds
                              (without alignments)
                              1540.951 Million cell updates/sec

Title: US-10-048-046-2_COPY_303_346
Perfect score: 266
Sequence: 1 TCIICQDLLHDCVSLQPCMH..........ACYSGWMERSSLCFTCRCPV 44

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	129	48.5	467	1	ICP0_VZVD	P09309 varicella-z
2	119	44.7	676	1	ICP0_HSVBJ	P29128 bovine herp
3	119	44.7	676	1	ICP0_HSVBK	P29836 bovine herp
4	110	41.4	610	1	YQDA_CAEEL	Q09268 caenorhabdi
5	108.5	40.8	485	1	RNF8_HUMAN	O76064 homo sapien
6	108.5	40.8	488	1	RNF8_MOUSE	Q8vc56 mus musculu
7	103.5	38.9	564	1	YUE5_CAEEL	P90859 caenorhabdi
8	99	37.2	740	1	BBAP_HUMAN	Q8tdb6 homo sapien
9	99	37.2	744	1	TRM3_HUMAN	075382 homo sapien
10	98	36.8	474	1	CBLC_HUMAN	Q9ulv8 homo sapien
11	98	36.8	532	1	ICP0_HSVEB	P28990 equine herp
12	97	36.5	154	1	RN11_HUMAN	Q9y3c5 homo sapien
13	97	36.5	154	1	RN11_MOUSE	Q9qyk7 mus musculu
14	97	36.5	825	1	ICP0_HSV2H	P28284 herpes simp
15	96	36.1	744	1	TRM3_MOUSE	Q9r1r2 mus musculu
16	96	36.1	744	1	TRM3_RAT	070277 rattus norv
17	96	36.1	906	1	CBL_HUMAN	P22681 homo sapien
18	96	36.1	913	1	CBL_MOUSE	P22682 mus musculu
19	96	36.1	982	1	CBLB_HUMAN	Q13191 homo sapien
20	95.5	35.9	643	1	AMF2_HUMAN	Q9ukv5 homo sapien
21	95.5	35.9	643	1	AMFR_MOUSE	Q9r049 mus musculu
22	95	35.7	410	1	ICP0_PRVIF	P29129 pseudorabie
23	95	35.7	775	1	ICP0_HSV11	P08393 herpes simp
24	94	35.3	324	1	BMI1_MOUSE	P25916 mus musculu
25	94	35.3	326	1	BMI1_FELCA	Q9tst0 felis silve
26	94	35.3	326	1	BMI1_HUMAN	P35226 homo sapien
27	93.5	35.2	407	1	RPF2_HUMAN	060858 homo sapien
28	92.5	34.8	387	1	RH18_SCHPO	074747 schizosacch
29	92	34.6	488	1	RN23_MOUSE	Q9esn2 mus musculu
30	92	34.6	518	1	RN23_HUMAN	Q9hcm9 homo sapien
31	91.5	34.4	605	1	PJA1_MOUSE	O55176 mus musculu
32	91.5	34.4	643	1	PJA1_HUMAN	Q8ng27 homo sapien
33	90.5	34.0	161	1	YOUD_CAEEL	P30631 caenorhabdi
34	89.5	33.6	377	1	RNG1_HUMAN	Q06587 homo sapien
35	88.5	33.3	443	1	UVSH_EMENI	Q02398 emericella
36	88.5	33.3	493	1	TRM5_HUMAN	Q9c035 homo sapien
37	88.5	33.3	501	1	UVS2_NEUCR	P33288 neurospora
38	88	33.1	342	1	MB18_MOUSE	P23798 mus musculu
39	88	33.1	344	1	MB18_HUMAN	P35227 homo sapien
40	87.5	32.9	337	1	PEXA_YEAST	Q05568 saccharomyc
41	87	32.7	481	1	RNF9_PANTR	Q7yr32 pan troglod
42	87	32.7	482	1	RNF9_HUMAN	Q9udy6 homo sapien
43	87	32.7	511	1	TRM7_HUMAN	Q9c029 homo sapien
44	87	32.7	1603	1	PSC_DROME	P35820 drosophila
45	85.5	32.1	115	1	YBR2_YEAST	P38239 saccharomyc

ALIGNMENTS

RESULT 1
ICP0_VZVD
ID   ICP0_VZVD   STANDARD;   PRT;   467 AA.
AC   P09309;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Trans-acting transcriptional protein ICP0.
GN   61.
OS   Varicella-zoster virus (strain Dumas) (VZV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=10338;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86306657; PubMed=3018124;
RA   Davison A.J., Scott J.E.;
RT   "The complete DNA sequence of varicella-zoster virus.";
RL   J. Gen. Virol. 67:1759-1816(1986).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=93059681; PubMed=1366099;
RA   Moriuchi H., Moriuchi M., Smith H.A., Straus S.E., Cohen J.I.;
RT   "Varicella-zoster virus open reading frame 61 protein is functionally
RT   homologous to herpes simplex virus type 1 ICP0.";
RL   J. Virol. 66:7303-7308(1992).
CC   -!- FUNCTION: REPRESSE THE EXPRESSION OF VIRAL IE, EARLY, AND LATE
CC   GENE PROMOTERS.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X04370; CAA27944.1; -.
DR   PIR; I27215; WZBE61.
DR   HSSP; P28990; 1CHC.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Transcription regulation; Trans-acting factor; Repressor; Zinc-finger;
KW   DNA-binding.
FT   ZN_FING   19   58   RING-TYPE.
SQ   SEQUENCE   467 AA; 50916 MW; 25EPA6977EA6994C CRC64;

Query Match   48.5%; Score 129; DB 1; Length 467;
Best Local Similarity   45.5%; Pred. No. 8.9e-08;
Matches   20; Conservative   2; Mismatches   22; Indels   0; Gaps   0;

A;Introns: 88/3; 132/2; 174/3; 222/3; 334/1; 493/3

```
Query Match          38.9%;      Score 103.5;  DB 2;   Length 564;
Best Local Similarity 41.9%;      Pred. No. 0.0022;
Matches   18; Conservative   6; Mismatches  18; Indels    1; Gaps    1;

Qy       2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
           |:::| :|| ||| ||||||:|
Db     335 CVVCWELLGTSRRLI-PCSHQFHDWCLMWWLAQDSSCPTCRCTI 376
```

RESULT 12
F71614
chromatinic RING finger DRING protein homolog PFB0440c - malaria parasite (Plasmodium fa
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 02-Sep-2000
A;Accession: F71614
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: F71614
A;Status: preliminary: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-568 <GAR>
A;Cross-references: GB:AE001395; GB:AE001362; NID:g3845184; PIDN:AAC71877.1; PID:g384518
C;Genetics:
A;Gene: PFB0440c
C;Superfamily: RING finger homology
F;210-260/Domain: RING finger homology <RRN>

```
Query Match          38.0%;      Score 101;  DB 2;   Length 568;
Best Local Similarity 37.8%;      Pred. No. 0.004;
Matches   17; Conservative  11; Mismatches  15; Indels    2; Gaps    1;

Qy       2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSL--CPTCRCPV 44
           ||| |:|:| ||| || ||||: | :::|:||| : |
Db     214 CPICLSLIHEPVTLNSCFHSFCWKCLATAIQKYSIDNCPSCRTKI 258
```

RESULT 13
B85092
hypothetical protein AT4g09100 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
A;Accession: B85092
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: B85092
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-132 <STO>
A;Cross-references: GB:NC_001268; NID:g7267553; PIDN:CAB78034.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g09100
A;Map position: 4

```
Query Match          37.8%;      Score 100.5;  DB 2;   Length 132;
Best Local Similarity 39.1%;      Pred. No. 0.0016;
Matches   18; Conservative   7; Mismatches  18; Indels    3; Gaps    1;

Qy       2 CIICQDLLHDCVSLQ--PCMHTFCAACYSGWMERSSLCPTCRCPV 44
           ||:::     :|| : |||||||: | ||||
Db      85 CVVCLNEFKDDETLRLVPPCVHVFHADCVDIWLSHSSTCPICRAKV 130
```

RESULT 14
T39456
zinc finger protein - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 23-Sep-2002
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Churcher, C.M.
submitted to the EMBL Data Library, June 1998
A;Reference number: Z21855
A;Accession: T39456
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-486 <LYN>
A;Cross-references: EMBL:AL023780; NID:e1295812; PIDN:CAA19328.1; GSPDB:GN00067; SPDB:SP
C;Genetics:
A;Gene: SPDB:SPBC14F5.10c
A;Map position: 2
C;Superfamily: zinc finger protein; RING finger homology
P;165-212/Domain: RING finger homology <RRN>

```
Query Match          37.8%;      Score 100.5;  DB 2;   Length 486;
Best Local Similarity 47.5%;      Pred. No. 0.004;
Matches   19; Conservative   5; Mismatches  15; Indels    1; Gaps    1;

Qy       2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCR 41
           | ::: :|| :|| |||||:| || :
Db     169 CQICFGMLYDPV-VSPCGHTFCGPCLMQALTQSPQCPTCR 207
```

RESULT 15
B96674
hypothetical protein F16G16.4 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
A;Accession: B96674
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: B96674
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-496 <STO>
A;Cross-references: GB:AE005173; NID:g6227002; PIDN:AAF06038.1; GSPDB:GN00141
C;Genetics:
A;Gene: F16G16.4
A;Map position: 1

```
Query Match          37.8%;      Score 100.5;  DB 2;   Length 496;
Best Local Similarity 43.2%;      Pred. No. 0.0041;
Matches   19; Conservative   5; Mismatches  19; Indels    1; Gaps    1;

Qy       1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
           |||||||||||:::
Db     327 TCIICREEMTSAKKKL-VCGHLFHVHCLRSWLERQNTCPTCRALV 369
```

Search completed: May 7, 2004, 14:50:08
Job time : 3.71251 secs

A;Molecule type: DNA
A;Residues: 1-551 <HUG>
A;Cross-references: EMBL:Z74755; NID:g1419784; PID:e252259; PID:g1419785; MIPS:YOL013c
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:HRD1
A;Cross-references: SGD:S0005373; MIPS:YOL013c
A;Map position: 15L
C;Keywords: transmembrane protein
C;Superfamily: RING finger homology
F;10-26/Domain: transmembrane #status predicted <TM1>
F;46-62/Domain: transmembrane #status predicted <TM2>
F;105-121/Domain: transmembrane #status predicted <TM3>
F;144-160/Domain: transmembrane #status predicted <TM4>
F;185-201/Domain: transmembrane #status predicted <TM5>
F;345-405/Domain: RING finger homology <RRN>

Query Match   40.2%;   Score 107;   DB 2;   Length 551;
Best Local Similarity   40.0%;   Pred. No. 0.00097;
Matches   22;   Conservative   2;   Mismatches   19;   Indels   12;   Gaps   1;

Qy   2 CIICQDLLHDCVSLQ----------PCMHTFCAACYSGWMERSSLCPTCRCPV 44
            ||||| || :| |||| :| ||| || |||
Db   349 CIICMDELIHSPNQQTWKNKNKKPKRLPCGHILHLSCLKNWMERSQTCPICRLPV 403

RESULT 8
F84591
hypothetical protein At2g20650 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: F84591
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: F84591
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <STO>
A;Cross-references: GB:AE002093; NID:g4512649; PIDN:AAD21704.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g20650
A;Map position: 2

Query Match   39.7%;   Score 105.5;   DB 2;   Length 276;
Best Local Similarity   42.9%;   Pred. No. 0.00086;
Matches   21;   Conservative   6;   Mismatches   15;   Indels   7;   Gaps   3;

Qy   2 CIICQ--DLLH---DCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
            |||| :||| |:| |||| :| ||| || |||
Db   226 CVICMTTIDLRHRINDCM-VTPCEHIFHSGCLQRWMDIKMECPTCRRPL 273

RESULT 9
T22687
hypothetical protein F55A11.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C;Accession: T22687
R;Kershaw, J.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z19600
A;Accession: T22687
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-610 <WIL>
A;Cross-references: EMBL:Z72511; PIDN:CAA96657.1; GSPDB:GN00023; CESP:F55A11.3
A;Experimental source: clone F55A11
C;Genetics:
A;Gene: CESP:F55A11.3

A;Map position: 5
A;Introns: 76/3; 221/1; 275/2; 338/1; 424/3
C;Superfamily: RING finger homology
F;288-338/Domain: RING finger homology <RRN>

Query Match   39.1%;   Score 104;   DB 2;   Length 610;
Best Local Similarity   41.3%;   Pred. No. 0.0021;
Matches   19;   Conservative   6;   Mismatches   19;   Indels   2;   Gaps   1;

Qy   1 TCIICQDLLHDCVSLQ--PCMHTFCAACYSGWMERSSLCPTCRCPV 44
            ||||| || : :|| :| :| ||| || |||
Db   291 TCIICRBMTVDASPKRLPCSHVPHAHCLRSWFQRQQTCPTCRTDI 336

RESULT 10
T16648
hypothetical protein R02E12.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Sep-2000
C;Accession: T16648
R;Leimbach, D.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid R02E12.
A;Reference number: Z18554
A;Accession: T16648
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-639 <LEI>
A;Cross-references: EMBL:U53337; NID:g1255833; PID:g1255835; PIDN:AAA96184.1; GSPDB:GN00(
A;Experimental source: strain Bristol N2; clone R02E12
C;Genetics:
A;Gene: CESP:R02E12.4
A;Map position: X
A;Introns: 63/3; 90/1; 136/2; 177/1; 220/2; 242/3; 294/2; 427/2; 464/3; 505/3; 580/1
C;Superfamily: RING finger homology
F;22-70/Domain: RING finger homology <RRN>

Query Match   39.1%;   Score 104;   DB 2;   Length 639;
Best Local Similarity   39.5%;   Pred. No. 0.0022;
Matches   17;   Conservative   6;   Mismatches   20;   Indels   0;   Gaps   0;

Qy   2 CIICQDLLHDCVSLQQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
            ||||| ||| : : |||| :| ||| || |||
Db   26 CHICFQVNHEPVTFLTCKHSICAGCAGRWLSSCSVCPMCRTRV 68

RESULT 11
T21423
hypothetical protein F26E4.11 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T21423; T21749
R;Lightning, J.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19419
A;Accession: T21423
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-564 <WIL>
A;Cross-references: EMBL:Z81070; PIDN:CAB03009.1; GSPDB:GN00019; CESP:F26E4.11
A;Experimental source: clone F26E4
R;Baynes, C.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19468
A;Accession: T21749
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-564 <WI2>
A;Cross-references: EMBL:Z81075; PIDN:CAB03049.1; GSPDB:GN00019; CESP:F26E4.11
A;Experimental source: clone F35C12
C;Genetics:
A;Gene: CESP:F26E4.11
A;Map position: 1

C;Keywords: DNA binding; transcription regulation; zinc finger
F;15-63/Domain: RING finger homology <RNG>
F;19-57/Region: zinc finger C3HC4 motif

```
Query Match             48.5%;  Score 129;  DB 1;  Length 467;
Best Local Similarity   45.5%;  Pred. No. 5.3e-06;
Matches   20;  Conservative   2;  Mismatches  22;  Indels   0;  Gaps   0;

Qy      1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
            ||||||    :|  |||:|                |||||
Db     18 TCTICMSTVSDLGKTMPCLHDFCFVCIRAWTSTSVQCPLCRCPV 61
```

RESULT 3
EDBB22
immediate-early protein IER2.9 - bovine herpesvirus 1 (strain K22)
N;Alternate names: early protein ER2.6; p135 protein
C;Species: bovine herpesvirus 1
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 17-Mar-2000
C;Accession: A38209
R;Wirth, U.V.; Fraefel, C.; Vogt, B.; Vlcek, C.; Paces, V.; Schwyzer, M.
J. Virol. 66, 2763-2772, 1992
A;Title: Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1 are 3' cotern
A;Reference number: A38209; MUID:92219360; PMID:1313901
A;Accession: A38209
A;Molecule type: DNA
A;Residues: 1-676 <WIR>
A;Cross-references: GB:M84464; NID:g330767; PIDN:AAA46061.1; PID:g330768
C;Superfamily: herpesvirus immediate-early protein IE110; RING finger homology
C;Keywords: DNA binding; immediate-early protein; transcription regulation; zinc finger
F;9-57/Domain: RING finger homology <RNG>
F;13-51/Region: zinc finger C3HC4 motif
F;284-331/Region: acidic

```
Query Match             44.7%;  Score 119;  DB 1;  Length 676;
Best Local Similarity   40.9%;  Pred. No. 6.9e-05;
Matches   18;  Conservative   5;  Mismatches   5;  Indels   0;  Gaps   0;

Qy      1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
            |||||| :  |||| |                |||| :||
Db     12 SCCICLDAITGAARALPCLHAFCLACIRRWLEGRPTCPLCKAPV 55
```

RESULT 4
EDBB23
immediate-early protein IER2.9 - bovine herpesvirus 1 (strain Jura)
N;Alternate names: early protein ER2.6; p135 protein
C;Species: bovine herpesvirus 1
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 17-Mar-2000
C;Accession: B38209
R;Wirth, U.V.; Fraefel, C.; Vogt, B.; Vlcek, C.; Paces, V.; Schwyzer, M.
J. Virol. 66, 2763-2772, 1992
A;Title: Immediate-early RNA 2.9 and early RNA 2.6 of bovine herpesvirus 1 are 3' cotern
A;Reference number: A38209; MUID:92219360; PMID:1313901
A;Accession: B38209
A;Molecule type: DNA
A;Residues: 1-676 <WIR>
A;Cross-references: GB:M84465; NID:g330769; PIDN:AAA46062.1; PID:g330770
C;Superfamily: herpesvirus immediate-early protein IE110; RING finger homology
C;Keywords: DNA binding; immediate-early protein; transcription regulation; zinc finger
F;9-57/Domain: RING finger homology <RNG>
F;13-51/Region: zinc finger C3HC4 motif
F;284-331/Region: acidic

```
Query Match             44.7%;  Score 119;  DB 1;  Length 676;
Best Local Similarity   40.9%;  Pred. No. 6.9e-05;
Matches   18;  Conservative   5;  Mismatches   5;  Indels   0;  Gaps   0;

Qy      1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
            |||||| :  |||| |                |||| :||
Db     12 SCCICLDAITGAARALPCLHAFCLACIRRWLEGRPTCPLCKAPV 55
```

RESULT 5
T15741
hypothetical protein C32D5.10 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Sep-2000
C;Accession: T15741
R;Miller, N.
submitted to the EMBL Data Library, July 1995
A;Description: The sequence of C. elegans cosmid C32D5.
A;Reference number: Z18396
A;Accession: T15741
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-621 <MIL>
A;Cross-references: EMBL:U23511; NID:g746463; PIDN:g746473; PIDN:AAC46798.1; CESP:C32D5.1
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:C32D5.10
A;Introns: 47/2; 141/2; 248/3; 292/3; 348/1; 383/1; 406/3; 429/1; 592/2; 610/1
C;Superfamily: RING finger homology
F;37-87/Domain: RING finger homology <RRN>

```
Query Match             41.4%;  Score 110;  DB 2;  Length 621;
Best Local Similarity   40.0%;  Pred. No. 0.00053;
Matches   18;  Conservative   7;  Mismatches  18;  Indels   2;  Gaps   1;

Qy      2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSS--LCPTCRCPV 44
            |:||:|||                          :|||||
Db     41 CSVCKNEIIDTTSLSDCCHEFCYDCIVGWLTKGSGPFCPMCKTPV 85
```

RESULT 6
T39702
probable peroxisome assembly protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C;Accession: T39702
R;Wood, V.; Skelton, J.; Churcher, C.M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, July 1999
A;Reference number: Z21870
A;Accession: T39702
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-306 <WOO>
A;Cross-references: EMBL:AL109652; PIDN:CAB51769.1; GSPDB:GN00067
A;Experimental source: strain 972h-; cosmid c17A3
C;Genetics:
A;Gene: Pi037
A;Map position: 2
A;Introns: 13/1
C;Superfamily: RING finger homology
F;252-299/Domain: RING finger homology <RNN>

```
Query Match             40.4%;  Score 107.5;  DB 2;  Length 306;
Best Local Similarity   40.0%;  Pred. No. 0.00058;
Matches   16;  Conservative   7;  Mismatches  16;  Indels   1;  Gaps   1;

Qy      2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCR 41
            |:|| : :|  ::| | :|:||:          :|||| ||
Db    256 CSLCMFIH-CPAATECGHIFCWSCINGWTSKKSECPLCR 294
```

RESULT 7
S66695
probable membrane protein YOL013c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein O2341
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 02-Sep-2000
C;Accession: S66695
R;Hughes, B.; Pohl, T.M.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66685
A;Accession: S66695

```
                  GenCore version 5.1.6
      Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May 7, 2004, 14:41:17 ; Search time 2.46251 Seconds
                                  (without alignments)
                          1718.743 Million cell updates/sec

Title:          US-10-048-046-2_COPY_303_346
Perfect score:  266
Sequence:       1 TCIICQDLLHDCVSLQPCMH.........ACYSGWMERSSLCPTCRCPV 44

scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283366 seqs, 9619526 residues

Total number of hits satisfying chosen parameters:        283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      PIR_78:*
            1:  Pir1:*
            2:  Pir2:*
            3:  Pir3:*
            4:  Pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                        SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	135	50.8	473	2	C96516	F16N3.15 [imported
2	129	48.5	467	1	WZBE61	gene 61 protein -
3	119	44.7	676	1	EDBB22	immediate-early pr
4	119	44.7	676	1	EDBB23	immediate-early pr
5	110	41.4	621	2	T15741	hypothetical prote
6	107.5	40.4	306	2	T39702	probable peroxisom
7	107	40.2	551	2	S66695	probable membrane
8	105.5	39.7	276	2	F84591	hypothetical prote
9	104	39.1	610	2	T22687	hypothetical prote
10	104	39.1	639	2	T16648	hypothetical prote
11	103.5	38.9	564	2	T21423	chromatinic RING f
12	101	38.0	568	2	F71614	hypothetical prote
13	100.5	37.8	132	2	B85092	zinc finger protei
14	100.5	37.8	486	2	T39456	hypothetical prote
15	100.5	37.8	496	2	B96674	hypothetical prote
16	100.5	37.8	1208	2	T00362	ring finger protei
17	100	37.6	222	2	JC4296	probable DNA repai
18	99.5	37.4	1375	2	T37672	spermatogenesis-re
19	98	36.8	230	2	JC7972	59K transcription
20	98	36.8	532	1	WZBEF5	probable transcrip
21	98	36.8	536	2	T42606	early protein EP0
22	97	36.5	410	2	A40505	immediate-early pr
23	97	36.5	825	1	EDBBXD	hypothetical prote
24	96.5	36.3	409	2	T25935	transforming prote
25	96	36.1	906	2	A43817	RNA binding-like p
26	95.5	35.9	245	2	T45552	immediate-early pr
27	95	35.7	315	1	D48560	immediate-early pr
28	95	35.7	775	1	EDBB11	hypothetical prote
29	94	35.3	118	2	T01883	hypothetical prote
30	94	35.3	324	2	A39523	transforming prote
31	94	35.3	326	2	I51694	transforming prote
32	94	35.3	326	2	I54339	proto-oncogene bmi
33	94	35.3	532	2	T49467	related to COP1-in
34	93.5	35.2	242	2	T48173	hypothetical prote
35	93.5	35.2	1331	2	T04938	hypothetical prote
36	93.5	35.2	1495	2	A85240	hypothetical prote
37	93.5	35.2	1495	2	T10649	hypothetical prote
38	92.5	34.8	387	2	T39653	probable DNA repai
39	92	34.6	273	2	T47692	testis-abundant fi
40	92	34.6	518	2	JC7387	hypothetical prote
41	91.5	34.4	302	2	C85092	probable RING zinc
42	91.5	34.4	327	2	D86474	hypothetical prote
43	91.5	34.4	345	2	D85092	hypothetical prote
44	91.5	34.4	836	2	T18460	hypothetical prote
45	91.5	34.4	2447	2	T16870	hypothetical prote

```
                        ALIGNMENTS

RESULT 1
C96516
F16N3.15 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: C96516
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
  Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C96516
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-473 <STO>
A;Cross-references: GB:AE005173; NID:g5668816; PIDN:AAD46042.1; GSPDB:GN00141
C;Genetics:
A;Gene: F16N3.15
A;Map position: 1

Query Match            50.8%;          Score 135;  DB 2;  Length 473;
Best Local Similarity  46.9%;          Pred. No. 1.3e-06;
Matches   23;  Conservative  5; Mismatches  15;  Indels  6;  Gaps  1;

Qy      2  CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSS-----LCPTCRCPV 44
           |||:||:||:|||| |:|| :|||| ||||| ||| |:|:| ||||||
Db    145  CSICLNIWHDVVTAAPCLHNFCNGCFSEWMRRSEEKHKGHVLCPQCRTTV 193

RESULT 2
WZBE61
gene 61 protein - human herpesvirus 3
C;Species: human herpesvirus 3, varicella-zoster virus
C;Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 17-Mar-2000
C;Accession: I27215
R;Davison, A.J.; Scott, J.E.
J. Gen. Virol. 67, 1759-1816, 1986
A;Title: The complete DNA sequence of varicella-zoster virus.
A;Reference number: A27345; MUID:86306657; PMID:3018124
A;Accession: I27215
A;Molecule type: DNA
A;Residues: 1-467 <DAV>
A;Cross-references: EMBL:X04370; NID:g55989; PIDN:CAA27944.1; PID:g60050
C;Genetics:
A;Gene: 61
C;Superfamily: varicella-zoster virus gene 61 protein; RING finger homology
```

```
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (358)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (494)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (504)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (505)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-808

Query Match             38.7%;  Score 103;  DB 11;  Length 505;
Best Local Similarity   39.5%;  Pred. No. 0.0035;
Matches   17;  Conservative    5;  Mismatches   21;  Indels    0;  Gaps     0;

Qy         2  CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
              ||||::.:    ||||||:  :  :|  ||||||| |||||
Db       291  CIICREEMVTGAKRLPCNHIFHTSCLRSWFQRQQTCPTCRMDV 333


RESULT 15
US-09-833-245-810
;   Sequence 810, Application US/09833245
;   Publication No. US20040010134A1
;   GENERAL INFORMATION:
;   APPLICANT: Human Genome Sciences, Inc.
;   TITLE OF INVENTION: Albumin Fusion Proteins
;   FILE REFERENCE: PF546PCT
;   CURRENT APPLICATION NUMBER: US/09/833,245
;   CURRENT FILING DATE: 2001-04-12
;   PRIOR APPLICATION NUMBER: 60/229, 358
;   PRIOR FILING DATE: 2000-04-12
;   PRIOR APPLICATION NUMBER: 60/256, 931
;   PRIOR FILING DATE: 2000-12-21
;   PRIOR APPLICATION NUMBER: 60/199, 384
;   PRIOR FILING DATE: 2000-04-25
;   NUMBER OF SEQ ID NOS: 2267
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 810
;   LENGTH: 617
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-833-245-810

Query Match             38.7%;  Score 103;  DB 11;  Length 617;
Best Local Similarity   39.5%;  Pred. No. 0.0041;
Matches   17;  Conservative    5;  Mismatches   21;  Indels    0;  Gaps     0;

Qy         2  CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
              ||||::.:    ||||||:  :  :|  ||||||| |||||
Db       291  CIICREEMVTGAKRLPCNHIFHTSCLRSWFQRQQTCPTCRMDV 333


Search completed: May  7, 2004, 15:06:51
Job time : 7.96938 secs
```

```
;    APPLICANT: Cao Yongwei
;    TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;    TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;    FILE REFERENCE: 38-21(53223)B
;    CURRENT APPLICATION NUMBER: US/10/424,599
;    CURRENT FILING DATE: 2003-04-28
;    NUMBER OF SEQ ID NOS: 285684
;    SEQ ID NO 235446
;    LENGTH: 216
;    TYPE: PRT
;    ORGANISM: Glycine max
;    FEATURE:
;    OTHER INFORMATION: Clone ID: PAT_MRT3847_54635C.1.pep
US-10-424-599-235446

Query Match          41.2%;     Score 109.5;  DB 12;  Length 216;
Best Local Similarity 39.1%;    Pred. No. 0.00033;
Matches  18;  Conservative  6;  Mismatches 19;  Indels 3;  Gaps 1;

Qy    2  CIICQDLLHD--CVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
         |:||:|:    :.||||| :  |||||
Db  139  CVVCLDAFHNAQWCRKLAACGHVFHRTCVDTWLLKVAACPTCRTPV 184

RESULT 11
US-10-424-599-250642
;    Sequence 250642, Application US/10424599
;    Publication No. US2004031072A1
;    GENERAL INFORMATION:
;    APPLICANT: La Rosa Thomas J
;    APPLICANT: Kovalic David K
;    APPLICANT: Zhou Yihua
;    APPLICANT: Cao Yongwei
;    TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;    TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;    FILE REFERENCE: 38-21(53223)B
;    CURRENT APPLICATION NUMBER: US/10/424,599
;    CURRENT FILING DATE: 2003-04-28
;    NUMBER OF SEQ ID NOS: 285684
;    SEQ ID NO 250642
;    LENGTH: 242
;    TYPE: PRT
;    ORGANISM: Glycine max
;    FEATURE:
;    OTHER INFORMATION: Clone ID: PAT_MRT3847_6835C.1.pep
US-10-424-599-250642

Query Match          40.8%;     Score 108.5;  DB 12;  Length 242;
Best Local Similarity 39.1%;    Pred. No. 0.00047;
Matches  18;  Conservative  6;  Mismatches 19;  Indels 3;  Gaps 1;

Qy    2  CIICQDLLHD--CVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
         |:||:|:    :.||||| :  |||||
Db  165  CVVCLDAFHNAQWCRKLAACGHVFHRRCVDTWLLKVAACPTCRTPV 210

RESULT 12
US-10-425-114-49876
;    Sequence 49876, Application US/10425114
;    Publication No. US2004034888A1
;    GENERAL INFORMATION:
;    APPLICANT: Liu, Jingdong
;    APPLICANT: Zhou, Yihua
;    APPLICANT: Kovalic, David K.
;    APPLICANT: Screen, Steven E
;    APPLICANT: Tabaska, Jack E
;    APPLICANT: Cao, Yongwei
;    TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;    TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;    FILE REFERENCE: 38-21(53313)B
;    CURRENT APPLICATION NUMBER: US/10/425,114
;    CURRENT FILING DATE: 2003-04-28
;    NUMBER OF SEQ ID NOS: 73128
```

```
;    SEQ ID NO 49876
;    LENGTH: 234
;    TYPE: PRT
;    ORGANISM: Glycine max
;    FEATURE:
;    OTHER INFORMATION: Clone ID: 700793076_FLI.pep
US-10-425-114-49876

Query Match          39.3%;     Score 104.5;  DB 12;  Length 234;
Best Local Similarity 44.2%;    Pred. No. 0.0012;
Matches  19;  Conservative  3;  Mismatches 20;  Indels 1;  Gaps 1;

Qy    2  CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
         |||:|      ||||||
Db  173  CAICQEKMQAPLLS-CKHMFCEECVSEWFERERTCPLCRALV 214

RESULT 13
US-10-424-599-154629
;    Sequence 154629, Application US/10424599
;    Publication No. US2004031072A1
;    GENERAL INFORMATION:
;    APPLICANT: La Rosa Thomas J
;    APPLICANT: Kovalic David K
;    APPLICANT: Zhou Yihua
;    APPLICANT: Cao Yongwei
;    TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;    TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;    FILE REFERENCE: 38-21(53223)B
;    CURRENT APPLICATION NUMBER: US/10/424,599
;    CURRENT FILING DATE: 2003-04-28
;    NUMBER OF SEQ ID NOS: 285684
;    SEQ ID NO 154629
;    LENGTH: 393
;    TYPE: PRT
;    ORGANISM: Glycine max
;    FEATURE:
;    NAME/KEY: unsure
;    LOCATION: (1)..(393)
;    OTHER INFORMATION: unsure at all Xaa locations
;    FEATURE:
;    OTHER INFORMATION: Clone ID: PAT_MRT3847_110651C.1.pep
US-10-424-599-154629

Query Match          39.3%;     Score 104.5;  DB 12;  Length 393;
Best Local Similarity 44.2%;    Pred. No. 0.0019;
Matches  19;  Conservative  3;  Mismatches 20;  Indels 1;  Gaps 1;

Qy    2  CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
         |||:|      ||||||
Db  332  CAICQEKMQAPLLS-CKHMFCEECVSEWFERERTCPLCRALV 373

RESULT 14
US-09-833-245-808
;    Sequence 808, Application US/09833245
;    Publication No. US20040101034A1
;    GENERAL INFORMATION:
;    APPLICANT: Human Genome Sciences, Inc.
;    TITLE OF INVENTION: Albumin Fusion Proteins
;    FILE REFERENCE: PF546PCT
;    CURRENT APPLICATION NUMBER: US/09/833,245
;    CURRENT FILING DATE: 2001-04-12
;    PRIOR APPLICATION NUMBER: 60/229, 358
;    PRIOR FILING DATE: 2000-04-12
;    PRIOR APPLICATION NUMBER: 60/256, 931
;    PRIOR FILING DATE: 2000-12-21
;    PRIOR APPLICATION NUMBER: 60/199, 384
;    PRIOR FILING DATE: 2000-04-25
;    NUMBER OF SEQ ID NOS: 2267
;    SOFTWARE: PatentIn Ver. 2.1
;    SEQ ID NO 808
;    LENGTH: 505
```

RESULT 6
US-10-001-885-150
; Sequence 150, Application US/10001885
; Publication No. US20040058319A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
; FILE REFERENCE: DEX-0279
; CURRENT APPLICATION NUMBER: US/10/001,885
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,061
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/253,257
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 167
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 150
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-885-150

Query Match   43.2%; Score 115; DB 12; Length 254;
Best Local Similarity 41.5%; Pred. No. 9.8e-05;
Matches   17; Conservative   9; Mismatches 15; Indels   0; Gaps   0;

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCR 41
         |||| :|| . .: :|::| |||||||||||||||||||||
Db 21 TCRLCSGYLIDATTVTECLHTFCRSCLVKYLEENNTCPTCR 61

RESULT 7
US-10-425-114-40033
; Sequence 40033, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 40033
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700962291_FLI.pep
US-10-425-114-40033

Query Match   42.3%; Score 112.5; DB 12; Length 183;
Best Local Similarity 46.5%; Pred. No. 0.00014;
Matches   20; Conservative   4; Mismatches 18; Indels   1; Gaps   1;

Qy   2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||:||:: .: :|:||::| |||||||||||||||||||||
Db 122 CAICQEKMHAPILLR-CKHIFCEDCVSEWFERERTCPLCRALV 163

RESULT 8

US-10-424-599-277916
; Sequence 277916, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 277916
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_92981C.1.pep
US-10-424-599-277916

Query Match   42.3%; Score 112.5; DB 12; Length 257;
Best Local Similarity 46.5%; Pred. No. 0.00018;
Matches   20; Conservative   4; Mismatches 18; Indels   1; Gaps   1;

Qy   2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||:||:: .: :|:||::| |||||||||||||||||||||
Db 196 CAICQEKMHAPILLR-CKHIFCEDCVSEWFERERTCPLCRALV 237

RESULT 9
US-10-425-114-51122
; Sequence 51122, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 51122
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700083617_FLI.pep
US-10-425-114-51122

Query Match   42.3%; Score 112.5; DB 12; Length 327;
Best Local Similarity 48.8%; Pred. No. 0.00023;
Matches   21; Conservative   3; Mismatches 18; Indels   1; Gaps   1;

Qy   2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||:||:: .: :|:||::| |||||||||||||||||||||
Db 266 CAICQEKMHVPVLLR-CKHIFCEDCVSEWFERERTCPLCRALV 307

RESULT 10
US-10-424-599-235446
; Sequence 235446, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua

NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 809
LENGTH: 426
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (414)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (415)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (420)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-809

Query Match   100.0%;   Score 266; DB 9; Length 426;
Best Local Similarity   100.0%;   Pred. No. 9.8e-21;
Matches   44; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          |||||||||||||||||||||||||||||||||||||||||||
Db 213 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 256

RESULT 3
US-09-780-525-2
; Sequence 2, Application US/09780525
; Patent No. US2002000 4223A1
; GENERAL INFORMATION:
APPLICANT: Bin-Bing Zhou
APPLICANT: Yuan Zhu
APPLICANT: Priya Chaturvedi
APPLICANT: Mark R. Hurle
APPLICANT: Xiaotong Li
TITLE OF INVENTION: FHAR1, A NEW RING FINGER PROTEIN
FILE REFERENCE: GP-70668-C1
CURRENT APPLICATION NUMBER: US/09/780,525
CURRENT FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 09/456,876
PRIOR FILING DATE: 1999-12-08
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 664
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-780-525-2

Query Match   100.0%;   Score 266; DB 9; Length 664;
Best Local Similarity   100.0%;   Pred. No. 1.4e-20;
Matches   44; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          |||||||||||||||||||||||||||||||||||||||||||
Db 303 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 346

RESULT 4
US-09-764-864-1270
; Sequence 1270, Application US/09764864
; Patent No. US2002013275A1
; GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PTZ23
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1270
LENGTH: 230
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (4)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (189)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (192)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (193)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (194)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (213)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (223)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (229)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (230)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1270

Query Match   66.2%;   Score 176; DB 9; Length 230;
Best Local Similarity   100.0%;   Pred. No. 2.6e-11;
Matches   28; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   17 PCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||||||||||||||||||||||||||
Db   5 PCMHTFCAACYSGWMERSSLCPTCRCPV 32

RESULT 5
US-09-919-039-285
; Sequence 285, Application US/0991 9039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
APPLICANT: Kaser, Matthew R.
TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
FILE REFERENCE: PA-0035 US
CURRENT APPLICATION NUMBER: US/09/919,039
CURRENT FILING DATE: 2002-09-09
PRIOR APPLICATION NUMBER: 60/222,113
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 401
SOFTWARE: PERL Program
SEQ ID NO 285
LENGTH: 247
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20030108871A1 1427459CD1
US-09-919-039-285

Query Match   43.2%;   Score 115; DB 10; Length 247;
Best Local Similarity   41.5%;   Pred. No. 9.6e-05;
Matches   17; Conservative   9; Mismatches   15; Indels   0; Gaps   0;

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCR 41
          ||| |:|||| :| |:|||:   ::|||||||||||||||
Db   21 TCRLCSGYLIDATTVTECLHTFCRSCLVKYLEENNTCPTCR 61

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May  7, 2004, 14:49:13 ; Search time 6.96938 Seconds
                                      (without alignments)
                                      1752.371 Million cell updates/sec

Title:           US-10-048-046-2_COPY_303_346
Perfect score:   266
Sequence:      1 TCIICQDLLHDCVSLQPCMH.........ACYSGWMERSSLCPTCRCPV 44

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters:        1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published Applications AA:*
                 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                 7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                 8: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
                10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
                11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
                12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
                14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
                15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
                16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result         %    Query
No.   Score  Match Length DB  ID                              Description
----  -----  ----- ------ --  --                              -----------
  1    266   100.0   349   15  US-10-108-260A-4516             Sequence 4516, Ap
  2    266   100.0   426    9  US-09-764-864-809               Sequence 809, App
  3    266   100.0   664    9  US-09-780-525-2                 Sequence 2, Appli
  4    176    66.2   230    9  US-09-764-864-1270              Sequence 1270, Ap
  5    115    43.2   247   10  US-09-919-039-285               Sequence 285, App
  6    115    43.2   254   12  US-10-001-885-150               Sequence 150, App
  7    112.5   42.3   183   12  US-10-425-114-40033             Sequence 40033, A
  8    112.5   42.3   257   12  US-10-424-599-277916            Sequence 277916,
  9    112.5   42.3   327   12  US-10-425-114-51122             Sequence 51122, A
 10    109.5   41.2   216   12  US-10-424-599-235446            Sequence 235446,
 11    108.5   40.8   242   12  US-10-424-599-250642            Sequence 250642,
 12    104.5   39.3   234   12  US-10-425-114-49876             Sequence 49876, A
 13    104.5   39.3   393   11  US-10-424-599-154629            Sequence 154629,
 14    104     38.7   505   11  US-09-833-245-808               Sequence 808, App
 15    103     38.7   617   11  US-09-833-245-810               Sequence 810, App
 16    103     38.7   617   12  US-10-470-360-17                Sequence 17, Appl
 17    100.5   38.7   185   12  US-10-425-114-40364             Sequence 40364, A
 18    99.5    37.4   165   12  US-10-424-599-170452            Sequence 170452,
 19    99.5    37.4   169   12  US-10-424-599-170451            Sequence 170451,
 20    99.5    37.4   176   12  US-10-425-114-49027             Sequence 49027, A
 21    99.5    37.4   206   12  US-10-424-599-180706            Sequence 180706,
 22    99      37.2   163    9  US-09-764-864-845               Sequence 845, App
 23    99      37.2   218    9  US-09-764-864-1300              Sequence 1300, Ap
 24    99      37.2   218   14  US-10-106-698-6052              Sequence 6052, Ap
 25    99      37.2   356   15  US-10-136-728-18                Sequence 18, Appl
 26    99      37.2   404   15  US-10-094-749-1721              Sequence 1721, Ap
 27    99      37.2   568   12  US-10-126-103-112               Sequence 112, App
 28    99      37.2   739    9  US-09-957-635-2                 Sequence 2, Appli
 29    98.5    37.0    48   14  US-10-283-500-14                Sequence 14, Appl
 30    98.5    37.0   113   12  US-10-424-599-158001            Sequence 158001,
 31    98.5    37.0   486   12  US-10-425-114-40759             Sequence 40759, A
 32    98.5    37.0   575   12  US-10-424-599-157327            Sequence 157327,
 33    98      36.8   131   12  US-10-264-049-2856              Sequence 2856, Ap
 34    98      36.8   158    9  US-09-764-864-1304              Sequence 1304, Ap
 35    98      36.8   230   15  US-10-295-027-702               Sequence 702, App
 36    98      36.8   439    9  US-09-764-864-849               Sequence 849, App
 37    98      36.8   439   15  US-10-264-049-3081              Sequence 3081, Ap
 38    98      36.8   474   15  US-10-157-031-251               Sequence 251, App
 39    97.5    36.7   331   12  US-10-425-114-37376             Sequence 37376, A
 40    97.5    36.7   353   12  US-10-424-599-187336            Sequence 187336,
 41    97.5    36.7   420   12  US-10-424-599-160966            Sequence 160966,
 42    97.5    36.7   424   12  US-10-425-114-70107             Sequence 70107, A
 43    97      36.5   154   12  US-10-221-625-105               Sequence 105, App
 44    97      36.5   207    9  US-09-925-302-638               Sequence 638, App
 45    97      36.5   207   12  US-09-925-302-638               Sequence 638, App
```

ALIGNMENTS

```
RESULT 1
US-10-108-260A-4516
; Sequence 4516, Application US/10108260A
; Publication No. US2004000556OA1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4516
;  LENGTH: 349
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-10-108-260A-4516

Query Match             100.0%;  Score 266;  DB 15;  Length 349;
Best Local Similarity   100.0%;  Pred. No. 8.2e-21;
Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
             ||||||||||||||||||||||||||||||||||||||||||||
Db       103 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 146

RESULT 2
US-09-764-864-809
; Sequence 809, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
```

```
            FILING DATE: 31-Mar-1998
            CLASSIFICATION: <Unknown>
        ATTORNEY/AGENT INFORMATION:
            NAME: Jackson Esq., David A.
            REGISTRATION NUMBER: 26,742
            REFERENCE/DOCKET NUMBER: 600-1-198 CIP 1
        TELECOMMUNICATION INFORMATION:
            TELEPHONE: 201-487-5800
            TELEFAX: 201-343-1684
    INFORMATION FOR SEQ ID NO: 12:
        SEQUENCE CHARACTERISTICS:
            LENGTH: 47 amino acids
            TYPE: amino acid
            STRANDEDNESS: single
            TOPOLOGY: linear
        MOLECULE TYPE: peptide
        HYPOTHETICAL: NO
        FRAGMENT TYPE: internal
        SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-052-089A-12

Query Match      36.1%;  Score 96;  DB 4;  Length 47;
Best Local Similarity  38.6%;  Pred. No. 0.00032;
Matches  17;  Conservative  6;  Mismatches  19;  Indels  2;  Gaps  2;

Qy    2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSL-CPTCRCPV 44
          ||| ||  |: |||||      |   :|   |||| -||  |||
Db    4 CKICAENDKD-VKIEPCGHLMCTSCLTSWQESEGQGCFCRCEI 46

RESULT 14
US-07-945-283-6
; Sequence 6, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
;     APPLICANT: Cheung, Andrew K.
;     APPLICANT: Wesley, Ronald D.
;     TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
;     TITLE OF INVENTION: Involving The EP0 and LLT Genes
;     NUMBER OF SEQUENCES: 7
;     CORRESPONDENCE ADDRESS:
;         ADDRESSEE: Curtis P. Ribando
;         STREET: 1815 No. 5352596th University Street
;         CITY: Peoria
;         STATE: IL
;         COUNTRY: USA
;         ZIP: 61604
;     COMPUTER READABLE FORM:
;         MEDIUM TYPE: Floppy disk
;         COMPUTER: IBM PC compatible
;         OPERATING SYSTEM: PC-DOS/MS-DOS
;         SOFTWARE: PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;         APPLICATION NUMBER: US/07/945,283
;         FILING DATE: 19920911
;         CLASSIFICATION: 424
;     ATTORNEY/AGENT INFORMATION:
;         NAME: Ribando, Curtis P
;         REGISTRATION NUMBER: 27976
;     TELECOMMUNICATION INFORMATION:
;         TELEPHONE: 309-685-4011 ext.513
;         TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 69 amino acids
;         TYPE: AMINO ACID
;         TOPOLOGY: linear
;     MOLECULE TYPE: peptide
;     FRAGMENT TYPE: internal
;     ORIGINAL SOURCE:
;         ORGANISM: Herpes simplex virus
US-07-945-283-6


Query Match      35.7%;  Score 95;  DB 1;  Length 69;
Best Local Similarity  39.0%;  Pred. No. 0.00058;
Matches  16;  Conservative  4;  Mismatches  19;  Indels  2;  Gaps  1;

Qy    2 CIICQDLL--HDCVSLQPCMHTFCAACYSGWMERSSLCPTC 40
          |||   ||  ||: ||||   | | :|  |||||  : |||
Db    6 CAVCTDEIAPHLRCDTFPCMHRFCIPCMKTWMQLRNTCPLC 46

RESULT 15
US-07-945-283-4
; Sequence 4, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
;     APPLICANT: Cheung, Andrew K.
;     APPLICANT: Wesley, Ronald D.
;     TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
;     TITLE OF INVENTION: Involving The EP0 and LLT Genes
;     NUMBER OF SEQUENCES: 7
;     CORRESPONDENCE ADDRESS:
;         ADDRESSEE: Curtis P. Ribando
;         STREET: 1815 No. 5352596th University Street
;         CITY: Peoria
;         STATE: IL
;         COUNTRY: USA
;         ZIP: 61604
;     COMPUTER READABLE FORM:
;         MEDIUM TYPE: Floppy disk
;         COMPUTER: IBM PC compatible
;         OPERATING SYSTEM: PC-DOS/MS-DOS
;         SOFTWARE: PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;         APPLICATION NUMBER: US/07/945,283
;         FILING DATE: 19920911
;         CLASSIFICATION: 424
;     ATTORNEY/AGENT INFORMATION:
;         NAME: Ribando, Curtis P
;         REGISTRATION NUMBER: 27976
;     TELECOMMUNICATION INFORMATION:
;         TELEPHONE: 309-685-4011 ext.513
;         TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 410 amino acids
;         TYPE: AMINO ACID
;         TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-07-945-283-4


Query Match      35.7%;  Score 95;  DB 1;  Length 410;
Best Local Similarity  39.5%;  Pred. No. 0.003;
Matches  17;  Conservative  2;  Mismatches  24;  Indels  0;  Gaps  0;

Qy    2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSLCPTCRCPV 44
          |||   || |:            :|  ||:|:  |||
Db   46 CPICLDVAATEAQTLPCMHKFCLDCIQRWTLTSTACPLCNARV 88

Search completed: May 7, 2004, 14:51:23
Job time : 3.92714 secs
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06639
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/250858
FILING DATE: 27-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/331394
FILING DATE: 28-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 897P2PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 48 amino acids
TYPE: amino acid
TOPOLOGY: linear
PCT-US95-06639-14

Query Match          37.0%;  Score  98.5;  DB 5;  Length 48;
Best Local Similarity  42.2%;  Pred. No. 0.00017;
Matches   19;  Conservative    5;  Mismatches  16;  Indels    5;  Gaps    2;

Qy    2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMER----SSLCPTCRC 42
        |||::|:|:||::|| |::      :  |         ::|||||||
Db    4 CKICAENDKD-VKIEPCGHLMCTSCLISWQESEGQGSSGCPPCRC 47


RESULT 11
US-07-945-283-5
; Sequence 5, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
APPLICANT: Cheung, Andrew K.
APPLICANT: Wesley, Ronald D.
TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
TITLE OF INVENTION: Involving The EP0 and LLT Genes
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis P. Ribando
STREET: 1815 No. 5352596th University Street
CITY: Peoria
STATE: IL
COUNTRY: USA
ZIP: 61604
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,283
FILING DATE: 19920911
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ribando, Curtis P
REGISTRATION NUMBER: 27976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 309-685-4011 ext.513
TELEFAX: 309-685-4128
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 67 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
```

```
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Pseudorabies virus
US-07-945-283-5

Query Match          36.5%;  Score  97;  DB 1;  Length 67;
Best Local Similarity  39.5%;  Pred. No. 0.00034;
Matches   17;  Conservative    3;  Mismatches  23;  Indels    0;  Gaps    0;

Qy    2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
        | ||:|:    ||:| |::    : :  |       :  :  |::
Db    6 CPICLDVAATEAQTLPCMHKFCLDCIQRWTLTSTACFLCKARV 48


RESULT 12
US-09-894-998A-47
; Sequence 47, Application US/09894998A
; Patent No. 6537555
; GENERAL INFORMATION:
APPLICANT: Hosken, Nancy Ann
APPLICANT: Craig H. Day
APPLICANT: Davin C. Dillon
APPLICANT: McGowan, Patrick
APPLICANT: Sleath, Paul R.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
FILE REFERENCE: 210121.538
CURRENT APPLICATION NUMBER: US/09/894,998A
CURRENT FILING DATE: 2001-06-28
NUMBER OF SEQ ID NOS: 64
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 47
LENGTH: 826
TYPE: PRT
ORGANISM: HSV-2
US-09-894-998A-47

Query Match          36.5%;  Score  97;  DB 4;  Length 826;
Best Local Similarity  35.6%;  Pred. No. 0.0034;
Matches   16;  Conservative    6;  Mismatches  21;  Indels    2;  Gaps    1;

Qy    2 CIICQDLLHDCVSLQ--PCMHTFCAACYSGWMERSSLCPTCRCPV 44
        | |:|:    ||:|||   |::||   |      : :||:|:|| |
Db  126 CAVCTDEIAPLRCQSFPCLHPFCIPCMKTWIPLRNTCPLCNTPV 170


RESULT 13
US-09-052-089A-12
; Sequence 12, Application US/09052089A
; Patent No. 6346605
; GENERAL INFORMATION:
APPLICANT: Lee, Soo Y.
APPLICANT: Choi, Yongwon
TITLE OF INVENTION: SIGNAL TRANSDUCER FOR THE TNF RECEPTOR SUPER
FAMILY, AND USES THEREOF
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/052,089A
APPLICATION NUMBER: US/09/052,089A
```

```
;     FILING DATE:  27-MAY-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/331394
;     FILING DATE:  28-OCT-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Dreger, Ginger R.
;     REGISTRATION NUMBER:  33,055
;     REFERENCE/DOCKET NUMBER:  89782
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  415/225-3216
;     TELEFAX:  415/952-9881
;     TELEX:  910/371-7168
; INFORMATION FOR SEQ ID NO:  14:
; SEQUENCE CHARACTERISTICS:
;   LENGTH:  48 amino acids
;   TYPE:  amino acid
;   TOPOLOGY:  linear
US-08-446-915-14

Query Match             37.0%;  Score 98.5;  DB 1;  Length 48;
Best Local Similarity   42.2%;  Pred. No. 0.00017;
Matches  19;  Conservative  5;  Mismatches 16;  Indels 5;  Gaps 2;

Qy  2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMER----SSLCPTCRC 42
       | ||| | |::||:||  :|:.:||| ||| |||
Db  4 CKICAENDKD-VKIEPCGHLMCTSCLTSWQESEGQGSSGCPFCRC 47

RSULT 8
US-08-744-139-14
; Sequence 14, Application US/08744139
; Patent No. 5869612
; GENERAL INFORMATION:
;   APPLICANT:  Goeddel, David V
;   APPLICANT:  Rothe, Mike
;   TITLE OF INVENTION:  TUMOR NECROSIS FACTOR RECPTOR-ASSOCIATED FACTORS
;   NUMBER OF SEQUENCES:  59
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Genentech, Inc.
;     STREET:  460 Point San Bruno Blvd
;     CITY:  South San Francisco
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  94080
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  WinPatin (Genentech)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/744,139
;     FILING DATE:  31-Oct-1996
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/250858
;     FILING DATE:  05/27/1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Dreger, Ginger R.
;     REGISTRATION NUMBER:  33,055
;     REFERENCE/DOCKET NUMBER:  P0897C1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  415/225-3216
;     TELEFAX:  415/952-9881
;     TELEX:  910/371-7168
; INFORMATION FOR SEQ ID NO:  14:
; SEQUENCE CHARACTERISTICS:
;   LENGTH:  48 amino acids
;   TYPE:  Amino Acid
;   TOPOLOGY:  Linear
US-08-744-139-14

Query Match             37.0%;  Score 98.5;  DB 2;  Length 48;
Best Local Similarity   42.2%;  Pred. No. 0.00017;
```

```
Matches  19;  Conservative  5;  Mismatches 16;  Indels 5;  Gaps 2;

Qy  2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMER----SSLCPTCRC 42
       | ||| | |::||:||  :|:.:||| ||| |||
Db  4 CKICAENDKD-VKIEPCGHLMCTSCLTSWQESEGQGSSGCPFCRC 47

RESULT 9
US-08-779-599-14
; Sequence 14, Application US/08775599
; Patent No. 6500922
; GENERAL INFORMATION:
;   APPLICANT:  Goeddel, David V.
;   APPLICANT:  Rothe, Mike
;   TITLE OF INVENTION:  Tumor Necrosis Factor Receptor-Associated Factors
;   NUMBER OF SEQUENCES:  59
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Genentech, Inc.
;     STREET:  460 Point San Bruno Blvd
;     CITY:  South San Francisco
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  94080
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  WinPatin (Genentech)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/779,599
;     FILING DATE:  07-Jan-1997
;     CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Dreger, Ginger R.
;     REGISTRATION NUMBER:  33,055
;     REFERENCE/DOCKET NUMBER:  P0897C2
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  415/225-3216
;     TELEFAX:  415/952-9881
;     TELEX:  910/371-7168
; INFORMATION FOR SEQ ID NO:  14:
; SEQUENCE CHARACTERISTICS:
;   LENGTH:  48 amino acids
;   TYPE:  Amino Acid
;   TOPOLOGY:  Linear
US-08-779-599-14

Query Match             37.0%;  Score 98.5;  DB 4;  Length 48;
Best Local Similarity   42.2%;  Pred. No. 0.00017;
Matches  19;  Conservative  5;  Mismatches 16;  Indels 5;  Gaps 2;

Qy  2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMER----SSLCPTCRC 42
       | ||| | |::||:||  :|:.:||| ||| |||
Db  4 CKICAENDKD-VKIEPCGHLMCTSCLTSWQESEGQGSSGCPFCRC 47

RESULT 10
PCT-US95-06639-14
; Sequence 14, Application PCT/TUS9506639
; GENERAL INFORMATION:
;   APPLICANT:  Genentech, Inc.
;   TITLE OF INVENTION:  Tumor Necrosis Factor Receptor-Associated Factors
;   NUMBER OF SEQUENCES:  66
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Genentech, Inc.
;     STREET:  460 Point San Bruno Blvd
;     CITY:  South San Francisco
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  94080
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  3.5 inch, 1.44 Mb disk
;     COMPUTER:  IBM PC compatible
```

```
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: polypeptide
US-09-628-133-25

Query Match           40.2%;  Score 107;   DB 4;  Length 551;
Best Local Similarity 40.0%;  Pred. No. 0.00018;
Matches    22;  Conservative  2;  Mismatches  19;  Indels  12;  Gaps  1;

Qy     2 CIICQDLLHDCVSLQ----------PCMHTFCAACYSGWMERSSLCPTCRCPV 44
         ||||           :    |     :| ||||| ||||| | ||||
Db   349 CIICMDELIHSPNQQTWKNKNKKPKRLPCGHILHLSCLKNWMERSQTCPICRLPV 403

RESULT 5
US-08-331-394-14
; Sequence 14, Application US/08331394
; Patent No. 5670319
; GENERAL INFORMATION:
; APPLICANT: Goeddel, David V.
; APPLICANT: Rothe, Mike
; TITLE OF INVENTION: Tumor Necrosis Factor
; TITLE OF INVENTION: Receptor-Associated Factors
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,394
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/250858
; FILING DATE: 27-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 897P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO:  14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-331-394-14

Query Match           37.0%;  Score 98.5;  DB 1;  Length 48;
Best Local Similarity 42.2%;  Pred. No. 0.00017;
Matches    19;  Conservative  5;  Mismatches  16;  Indels  16;  Gaps  2;

Qy     2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMER----SSLCPTCRC 42
         ||||        |::||           |·|       |||||
Db     4 CKICAENDKD-VKIEPCGHLMCTSCLTSWQESEGGQGSSGCPFCRC 47

RESULT 6
US-08-250-858-14
; Sequence 14, Application US/08250858
; Patent No. 5708142
; GENERAL INFORMATION:
; APPLICANT: Goeddel, David V.
; APPLICANT: Rothe, Mike
; TITLE OF INVENTION: Tumor Necrosis Factor Receptor-Associated Factors
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/250,858
; FILING DATE: 27-May-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 897.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO:  14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-250-858-14

Query Match           37.0%;  Score 98.5;  DB 1;  Length 48;
Best Local Similarity 42.2%;  Pred. No. 0.00017;
Matches    19;  Conservative  5;  Mismatches  16;  Indels  16;  Gaps  2;

Qy     2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMER----SSLCPTCRC 42
         ||||        |::||           |·|       |||||
Db     4 CKICAENDKD-VKIEPCGHLMCTSCLTSWQESEGGQGSSGCPFCRC 47

RESULT 7
US-08-446-915-14
; Sequence 14, Application US/08446915
; Patent No. 5741667
; GENERAL INFORMATION:
; APPLICANT: Goeddel, David V.
; APPLICANT: Rothe, Mike
; TITLE OF INVENTION: Tumor Necrosis Factor Receptor-Associated Factors
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,915
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/250858
```

RESULT 2
US-08-699-103B-25
; Sequence 25, Application US/08699103B
; Patent No. 6107462
; GENERAL INFORMATION:
;; APPLICANT: Rine, Jasper D.
;; APPLICANT: Hampton, Randolph
;; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
;; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
;; NUMBER OF SEQUENCES: 25
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fish & Richardson P.C.
;; STREET: 2200 Sand Hill Road, Suite 100
;; CITY: Menlo Park
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94025
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSEQ for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/699,103B
;; FILING DATE: 16-AUG-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 60/002,581
;; FILING DATE: 17-AUG-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Green, Grant D.
;; REGISTRATION NUMBER: 31,259
;; REFERENCE/DOCKET NUMBER: 09272/005001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650/322-5070
;; TELEFAX: 650/854-0875
; INFORMATION FOR SEQ ID NO: 25:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 551 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: polypeptide
US-08-699-103B-25

Query Match   40.2%; Score 107; DB 3; Length 551;
Best Local Similarity 40.0%; Pred. No. 0.00018;
Matches 22; Conservative 2; Mismatches 19; Indels 12; Gaps 1;

Qy   2 CIICQDLLHDCVSLQ----------PCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||| | | ||||||||
Db 349 CIICMDELIHSPNQQTWKNKKPKRLPCGHILHLSCLKQNWMERSQTCPICRLPV 403

RESULT 3
US-09-229-059-25
; Sequence 25, Application US/09229059
; Patent No. 6333172
; GENERAL INFORMATION:
;; APPLICANT: Rine, Jasper D.
;; APPLICANT: Hampton, Randolph
;; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
;; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
;; NUMBER OF SEQUENCES: 25
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fish & Richardson P.C.
;; STREET: 2200 Sand Hill Road, Suite 100
;; CITY: Menlo Park
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94025
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSEQ for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/229,059
;; FILING DATE:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/699,103
;; FILING DATE: 16-AUG-1996
;; APPLICATION NUMBER: 60/002,581
;; FILING DATE: 17-AUG-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Green, Grant D.
;; REGISTRATION NUMBER: 31,259
;; REFERENCE/DOCKET NUMBER: 09272/005001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650/322-5070
;; TELEFAX: 650/854-0875
; INFORMATION FOR SEQ ID NO: 25:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 551 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: polypeptide
US-09-229-059-25

Query Match   40.2%; Score 107; DB 4; Length 551;
Best Local Similarity 40.0%; Pred. No. 0.00018;
Matches 22; Conservative 2; Mismatches 19; Indels 12; Gaps 1;

Qy   2 CIICQDLLHDCVSLQ----------PCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||| | | ||||||||
Db 349 CIICMDELIHSPNQQTWKNKKPKRLPCGHILHLSCLKQNWMERSQTCPICRLPV 403

RESULT 4
US-09-628-133-25
; Sequence 25, Application US/09628133
; Patent No. 6531292
; GENERAL INFORMATION:
;; APPLICANT: Rine, Jasper D.
;; APPLICANT: Hampton, Randolph
;; TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING
;; TITLE OF INVENTION: CHOLESTEROL SYNTHESIS
;; NUMBER OF SEQUENCES: 25
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fish & Richardson P.C.
;; STREET: 2200 Sand Hill Road, Suite 100
;; CITY: Menlo Park
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94025
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSEQ for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/628,133
;; FILING DATE:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/699,103
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Green, Grant D.
;; REGISTRATION NUMBER: 31,259
;; REFERENCE/DOCKET NUMBER: 09272/005001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650/322-5070
;; TELEFAX: 650/854-0875
; INFORMATION FOR SEQ ID NO: 25:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 551 amino acids

```
                    GenCore version 5.1.6
         Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May  7, 2004, 14:42:23 ; Search time 2.92714 Seconds
                                     (without alignments)
                                     776.028 Million cell updates/sec

Title:       US-10-048-046-2_COPY_303_346
Perfect score:  266
Sequence:    1 TCIICQDLIHDCVSLQPCMH..........ACYSGWMERSSLCPTCRCPV 44

Scoring table:  BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:    389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:    389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued Patents AA:*
             1: /cgn2_6/ptcdata/2/iaa/5A_COMB.pep:*
             2: /cgn2_6/ptcdata/2/iaa/5B_COMB.pep:*
             3: /cgn2_6/ptcdata/2/iaa/6A_COMB.pep:*
             4: /cgn2_6/ptcdata/2/iaa/6B_COMB.pep:*
             5: /cgn2_6/ptcdata/2/iaa/PCTUS_COMB.pep:*
             6: /cgn2_6/ptcdata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	Query Match %	Match Length	DB	ID	Description
1	129	48.5	67	1	US-07-945-283-7	Sequence 7, Appli
2	107	40.2	551	3	US-08-699-103B-25	Sequence 25, Appl
3	107	40.2	551	4	US-09-229-059-25	Sequence 25, Appl
4	107	40.2	551	1	US-09-628-133-25	Sequence 25, Appl
5	98.5	37.0	48	1	US-08-331-394-14	Sequence 14, Appl
6	98.5	37.0	48	1	US-08-250-858-14	Sequence 14, Appl
7	98.5	37.0	48	2	US-08-446-915-14	Sequence 14, Appl
8	98.5	37.0	48	4	US-08-744-139-14	Sequence 14, Appl
9	98.5	37.0	48	4	US-08-779-599-14	Sequence 14, Appl
10	98.5	37.0	48	5	PCT-US95-06639-14	Sequence 14, Appl
11	97	36.5	67	1	US-07-945-283-5	Sequence 5, Appl
12	97	36.5	826	4	US-09-894-998A-47	Sequence 47, Appl
13	96	36.1	47	4	US-09-052-089A-12	Sequence 12, Appl
14	95	35.7	69	1	US-07-945-283-6	Sequence 6, Appli
15	95	35.7	410	1	US-07-945-283-4	Sequence 4, Appl
16	94	35.3	326	2	US-08-786-606-8	Sequence 8, Appli
17	92.5	34.8	336	3	US-08-972-719-2	Sequence 2, Appl
18	92.5	34.8	336	.	US-08-753-038-2	Sequence 2, Appl
19	92.5	34.8	336	4	US-09-590-175-2	Sequence 2, Appl
20	91.5	34.4	223	4	US-09-327-983-5	Sequence 5, Appli
21	91.5	34.4	395	4	US-08-841-349-9	Sequence 9, Appli
22	91.5	34.4	395	4	US-08-431-184A-9	Sequence 9, Appli
23	91	34.2	245	2	US-08-897-340-36	Sequence 36, Appl
24	91	34.2	245	3	US-09-252-329-36	Sequence 36, Appl
25	91	34.2	317	4	US-09-921-099A-8	Sequence 8, Appli
26	90	33.8	305	4	US-09-599-360B-114	Sequence 114, App
27	89.5	33.6	41	1	US-08-480-784-84	Sequence 84, Appl

28	89.5	33.6	41	1	US-08-483-553-84	Sequence 84, Appl
29	89.5	33.6	41	1	US-08-487-002-84	Sequence 84, Appl
30	89.5	33.6	41	1	US-08-483-554B-84	Sequence 84, Appl
31	89.5	33.6	41	1	US-08-488-011B-84	Sequence 84, Appl
32	89.5	33.6	41	3	US-08-850-727-84	Sequence 84, Appl
33	89.5	33.6	41	5	PCT-US95-10202-84	Sequence 84, Appl
34	89.5	33.6	41	5	PCT-US95-10203-84	Sequence 84, Appl
35	89.5	33.6	41	5	PCT-US95-10220-84	Sequence 84, Appl
36	89.5	33.6	46	1	US-08-331-394-11	Sequence 11, Appl
37	89.5	33.6	46	1	US-08-250-858-11	Sequence 11, Appl
38	89.5	33.6	46	2	US-08-446-915-11	Sequence 11, Appl
39	89.5	33.6	46	4	US-08-744-139-11	Sequence 11, Appl
40	89.5	33.6	46	4	US-08-779-599-11	Sequence 11, Appl
41	89.5	33.6	46	5	PCT-US95-06639-11	Sequence 11, Appl
42	89.5	33.6	47	2	US-08-691-814B-13	Sequence 13, Appl
43	89.5	33.6	48	4	US-09-052-089A-13	Sequence 13, Appl
44	88.5	33.3	44	8	US-08-331-394-8	Sequence 8, Appli
45	88.5	33.3	44	1	US-08-250-858-8	Sequence 8, Appli

```
                              ALIGNMENTS

RESULT 1
US-07-945-283-7
; Sequence 7, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
;  APPLICANT:  Cheung, Andrew K.
;  APPLICANT:  Wesley, Ronald D.
;  TITLE OF INVENTION:  Pseudorabies Virus Deletion Mutants
;  TITLE OF INVENTION:  Involving The EP0 and LLT Genes
;  NUMBER OF SEQUENCES:  7
;  CORRESPONDENCE ADDRESS:
;   ADDRESSE:  Curtis P. Ribando
;   STREET:  1815 No. 5352596th University Street
;   CITY:  Peoria
;   STATE:  IL
;   COUNTRY:  USA
;   ZIP:  61604
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE:  Floppy disk
;   COMPUTER:  IBM PC compatible
;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/07/945,283
;   FILING DATE:  19920911
;   CLASSIFICATION:  424
;  ATTORNEY/AGENT INFORMATION:
;   NAME:  Ribando, Curtis P
;   REGISTRATION NUMBER:  27976
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  309-685-4011 ext.513
;   TELEFAX:  309-685-4128
;  INFORMATION FOR SEQ ID NO:  7:
;   SEQUENCE CHARACTERISTICS:
;    LENGTH:  67 amino acids
;    TYPE:  AMINO ACID
;    TOPOLOGY:  linear
;   MOLECULE TYPE:  peptide
;   FRAGMENT TYPE:  internal
;   ORIGINAL SOURCE:
;    ORGANISM:  Varicella-zoster virus
; US-07-945-283-7

  Query Match             48.5%;  Score 129;  DB 1;  Length 67;
  Best Local Similarity   45.5%;  Pred. No. 9.5e-08;
  Matches   20;  Conservative   2;  Mismatches  22;  Indels   0;  Gaps   0;

Qy    1 TCIICQDLIHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||  ||| |||:|: |  :  ||:|||||||||||
Db    5 TCTICMSTVSDLGKTMPCLHDFCFVCIRAWTSTSVQCPLCRCPV 48
```

```
PR   05-OCT-1999;   99US-0157753P.
PR   06-OCT-1999;   99US-0157865P.
PR   07-OCT-1999;   99US-0158029P.
PR   08-OCT-1999;   99US-0158232P.
PR   12-OCT-1999;   99US-0158369P.
PR   13-OCT-1999;   99US-0159293P.
PR   13-OCT-1999;   99US-0159294P.
PR   13-OCT-1999;   99US-0159295P.
PR   14-OCT-1999;   99US-0159329P.
PR   14-OCT-1999;   99US-0159310P.
PR   14-OCT-1999;   99US-0159331P.
PR   14-OCT-1999;   99US-0159637P.
PR   14-OCT-1999;   99US-0159638P.
PR   18-OCT-1999;   99US-0159584P.
PR   21-OCT-1999;   99US-0160741P.
PR   21-OCT-1999;   99US-0160767P.
PR   21-OCT-1999;   99US-0160768P.
PR   21-OCT-1999;   99US-0160770P.
PR   21-OCT-1999;   99US-0160814P.
PR   21-OCT-1999;   99US-0160815P.
PR   22-OCT-1999;   99US-0160980P.
PR   22-OCT-1999;   99US-0160981P.
PR   22-OCT-1999;   99US-0160989P.
PR   22-OCT-1999;   99US-0161404P.
PR   25-OCT-1999;   99US-0161405P.
PR   25-OCT-1999;   99US-0161406P.
PR   26-OCT-1999;   99US-0161359P.
PR   26-OCT-1999;   99US-0161360P.
PR   26-OCT-1999;   99US-0161361P.
PR   28-OCT-1999;   99US-0161920P.
PR   28-OCT-1999;   99US-0161992P.
PR   28-OCT-1999;   99US-0161993P.
PR   29-OCT-1999;   99US-0162142P.
```

```
Query Match         50.8%;   Score 135;  DB 3;   Length 453;
Best Local Similarity   46.9%;   Pred. No. 1.2e-06;
Matches   23;  Conservative   5;  Mismatches   15;  Indels   6;  Gaps   1;

Qy     2  CIICCDLLHDCVSLQPCMHTFCAACYSGWMERSS------LCPTCRCPV  44
             |||  :: ||  || :|| :||::  |:|:           |||:|||  |
Db   125  CSICLNIWHDVVTAAPCLHNFCNGCFSEWMRRSEEKHKHVLCPQCRTTV  173
```

Search completed: May 7, 2004, 14:45:48
Job time : 11.2637 secs

PR 05-MAR-1999, 99US-0123180P.
PR 09-MAR-1999, 99US-0123548P.
PR 23-MAR-1999, 99US-0125788P.
PR 25-MAR-1999, 99US-0126264P.
PR 29-MAR-1999, 99US-0126785P.
PR 01-APR-1999, 99US-0127462P.
PR 06-APR-1999, 99US-0128234P.
PR 08-APR-1999, 99US-0128714P.
PR 16-APR-1999, 99US-0129845P.
PR 19-APR-1999, 99US-0130077P.
PR 21-APR-1999, 99US-0130449P.
PR 23-APR-1999, 99US-0130510P.
PR 23-APR-1999, 99US-0130891P.
PR 28-APR-1999, 99US-0131449P.
PR 30-APR-1999, 99US-0132048P.
PR 30-APR-1999, 99US-0132407P.
PR 04-MAY-1999, 99US-0132484P.
PR 05-MAY-1999, 99US-0132485P.
PR 06-MAY-1999, 99US-0132486P.
PR 06-MAY-1999, 99US-0132487P.
PR 07-MAY-1999, 99US-0132863P.
PR 11-MAY-1999, 99US-0134256P.
PR 14-MAY-1999, 99US-0134218P.
PR 14-MAY-1999, 99US-0134219P.
PR 14-MAY-1999, 99US-0134221P.
PR 18-MAY-1999, 99US-0134370P.
PR 18-MAY-1999, 99US-0134768P.
PR 19-MAY-1999, 99US-0134941P.
PR 20-MAY-1999, 99US-0135124P.
PR 21-MAY-1999, 99US-0135353P.
PR 24-MAY-1999, 99US-0135629P.
PR 25-MAY-1999, 99US-0136021P.
PR 27-MAY-1999, 99US-0136392P.
PR 28-MAY-1999, 99US-0136782P.
PR 01-JUN-1999, 99US-0137222P.
PR 03-JUN-1999, 99US-0137528P.
PR 04-JUN-1999, 99US-0137502P.
PR 07-JUN-1999, 99US-0137724P.
PR 08-JUN-1999, 99US-0138094P.
PR 08-JUN-1999, 99US-0138540P.
PR 10-JUN-1999, 99US-0138847P.
PR 14-JUN-1999, 99US-0139119P.
PR 16-JUN-1999, 99US-0139452P.
PR 16-JUN-1999, 99US-0139453P.
PR 17-JUN-1999, 99US-0139492P.
PR 18-JUN-1999, 99US-0139454P.
PR 18-JUN-1999, 99US-0139455P.
PR 18-JUN-1999, 99US-0139456P.
PR 18-JUN-1999, 99US-0139457P.
PR 18-JUN-1999, 99US-0139458P.
PR 18-JUN-1999, 99US-0139459P.
PR 18-JUN-1999, 99US-0139460P.
PR 18-JUN-1999, 99US-0139461P.
PR 18-JUN-1999, 99US-0139462P.
PR 18-JUN-1999, 99US-0139463P.
PR 18-JUN-1999, 99US-0139750P.
PR 18-JUN-1999, 99US-0139763P.
PR 21-JUN-1999, 99US-0139817P.
PR 22-JUN-1999, 99US-0139899P.
PR 23-JUN-1999, 99US-0140353P.
PR 24-JUN-1999, 99US-0140354P.
PR 24-JUN-1999, 99US-0140695P.
PR 28-JUN-1999, 99US-0140823P.
PR 29-JUN-1999, 99US-0140991P.
PR 30-JUN-1999, 99US-0141287P.
PR 01-JUL-1999, 99US-0141842P.
PR 01-JUL-1999, 99US-0142154P.
PR 02-JUL-1999, 99US-0142055P.
PR 06-JUL-1999, 99US-0142390P.
PR 08-JUL-1999, 99US-0142803P.
PR 09-JUL-1999, 99US-0142920P.
PR 12-JUL-1999, 99US-0142977P.
PR 13-JUL-1999, 99US-0143542P.

PR 14-JUL-1999, 99US-0143624P.
PR 15-JUL-1999, 99US-0144005P.
PR 16-JUL-1999, 99US-0144005P.
PR 16-JUL-1999, 99US-0144085P.
PR 19-JUL-1999, 99US-0144086P.
PR 19-JUL-1999, 99US-0144325P.
PR 19-JUL-1999, 99US-0144331P.
PR 19-JUL-1999, 99US-0144332P.
PR 19-JUL-1999, 99US-0144333P.
PR 19-JUL-1999, 99US-0144334P.
PR 20-JUL-1999, 99US-0144335P.
PR 20-JUL-1999, 99US-0144352P.
PR 20-JUL-1999, 99US-0144632P.
PR 20-JUL-1999, 99US-0144884P.
PR 21-JUL-1999, 99US-0144814P.
PR 21-JUL-1999, 99US-0145006P.
PR 22-JUL-1999, 99US-0145088P.
PR 22-JUL-1999, 99US-0145085P.
PR 22-JUL-1999, 99US-0145087P.
PR 23-JUL-1999, 99US-0145089P.
PR 23-JUL-1999, 99US-0145145P.
PR 23-JUL-1999, 99US-0145192P.
PR 26-JUL-1999, 99US-0145218P.
PR 27-JUL-1999, 99US-0145224P.
PR 27-JUL-1999, 99US-0145276P.
PR 27-JUL-1999, 99US-0145913P.
PR 28-JUL-1999, 99US-0145918P.
PR 02-AUG-1999, 99US-0145919P.
PR 02-AUG-1999, 99US-0145951P.
PR 03-AUG-1999, 99US-0146386P.
PR 04-AUG-1999, 99US-0146388P.
PR 04-AUG-1999, 99US-0146389P.
PR 05-AUG-1999, 99US-0147038P.
PR 06-AUG-1999, 99US-0147204P.
PR 09-AUG-1999, 99US-0147302P.
PR 09-AUG-1999, 99US-0147192P.
PR 10-AUG-1999, 99US-0147260P.
PR 11-AUG-1999, 99US-0147303P.
PR 12-AUG-1999, 99US-0147416P.
PR 13-AUG-1999, 99US-0147493P.
PR 13-AUG-1999, 99US-0147935P.
PR 16-AUG-1999, 99US-0148171P.
PR 17-AUG-1999, 99US-0148319P.
PR 18-AUG-1999, 99US-0148341P.
PR 20-AUG-1999, 99US-0148565P.
PR 20-AUG-1999, 99US-0148684P.
PR 23-AUG-1999, 99US-0149368P.
PR 25-AUG-1999, 99US-0149175P.
PR 26-AUG-1999, 99US-0149426P.
PR 27-AUG-1999, 99US-0149722P.
PR 27-AUG-1999, 99US-0149723P.
PR 30-AUG-1999, 99US-0149929P.
PR 31-AUG-1999, 99US-0149902P.
PR 01-SEP-1999, 99US-0149930P.
PR 07-SEP-1999, 99US-0150566P.
PR 10-SEP-1999, 99US-0150884P.
PR 13-SEP-1999, 99US-0151065P.
PR 15-SEP-1999, 99US-0151066P.
PR 16-SEP-1999, 99US-0151080P.
PR 20-SEP-1999, 99US-0151303P.
PR 22-SEP-1999, 99US-0151438P.
PR 23-SEP-1999, 99US-0151930P.
PR 24-SEP-1999, 99US-0152363P.
PR 28-SEP-1999, 99US-0153070P.
PR 29-SEP-1999, 99US-0153758P.
PR 04-OCT-1999, 99US-0154018P.
                  99US-0154039P.
                  99US-0154779P.
                  99US-0155139P.
                  99US-0155486P.
                  99US-0155659P.
                  99US-0156458P.
                  99US-0156596P.
                  99US-0157117P.

```
Qy          2  CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSS-----LCPTCRCPV 44
               | |::| | | | |:| | | |   | |:| |||| | ||||| ||
Db         22  CSICLNIWHDVVTAAPCLHNPCNGCFSEWMRRSEEKHKGHVLCPQCRTTV 70


RESULT 11
AAG37660
ID   AAG37660 standard; protein; 350 AA.
XX
AC   AAG37660;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 46344.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999;  99US-0121825P.
PR   05-MAR-1999;  99US-0123180P.
PR   09-MAR-1999;  99US-0123548P.
PR   23-MAR-1999;  99US-0125788P.
PR   25-MAR-1999;  99US-0126264P.
PR   29-MAR-1999;  99US-0126785P.
PR   01-APR-1999;  99US-0127462P.
PR   06-APR-1999;  99US-0128234P.
PR   08-APR-1999;  99US-0128714P.
PR   16-APR-1999;  99US-0129845P.
PR   19-APR-1999;  99US-0130077P.
PR   21-APR-1999;  99US-0130449P.
PR   23-APR-1999;  99US-0130510P.
PR   23-APR-1999;  99US-0130891P.
PR   28-APR-1999;  99US-0131449P.
PR   30-APR-1999;  99US-0132048P.
PR   30-APR-1999;  99US-0132407P.
PR   04-MAY-1999;  99US-0132484P.
PR   05-MAY-1999;  99US-0132485P.
PR   06-MAY-1999;  99US-0132486P.
PR   06-MAY-1999;  99US-0132487P.
PR   07-MAY-1999;  99US-0132863P.
PR   11-MAY-1999;  99US-0132392P.
PR   14-MAY-1999;  99US-0134256P.
PR   14-MAY-1999;  99US-0134218P.
PR   14-MAY-1999;  99US-0134219P.
PR   14-MAY-1999;  99US-0134221P.
PR   14-MAY-1999;  99US-0134370P.
PR   18-MAY-1999;  99US-0134768P.
PR   19-MAY-1999;  99US-0134941P.
PR   20-MAY-1999;  99US-0135124P.
PR   21-MAY-1999;  99US-0135353P.
PR   24-MAY-1999;  99US-0135629P.
PR   25-MAY-1999;  99US-0136021P.
PR   27-MAY-1999;  99US-0136392P.
PR   28-MAY-1999;  99US-0136782P.
PR   01-JUN-1999;  99US-0137222P.
PR   03-JUN-1999;  99US-0137528P.
PR   04-JUN-1999;  99US-0137502P.
PR   07-JUN-1999;  99US-0137724P.
PR   08-JUN-1999;  99US-0138094P.
PR   10-JUN-1999;  99US-0138540P.
PR   10-JUN-1999;  99US-0138847P.
PR   14-JUN-1999;  99US-0139119P.
PR   16-JUN-1999;  99US-0139452P.
PR   16-JUN-1999;  99US-0139453P.
PR   17-JUN-1999;  99US-0139492P.
PR   18-JUN-1999;  99US-0139454P.
PR   18-JUN-1999;  99US-0139455P.
PR   18-JUN-1999;  99US-0139456P.
PR   18-JUN-1999;  99US-0139457P.
PR   18-JUN-1999;  99US-0139458P.
PR   18-JUN-1999;  99US-0139459P.
PR   18-JUN-1999;  99US-0139460P.
PR   18-JUN-1999;  99US-0139461P.
PR   18-JUN-1999;  99US-0139462P.
PR   18-JUN-1999;  99US-0139463P.
PR   18-JUN-1999;  99US-0139750P.
PR   18-JUN-1999;  99US-0139763P.
PR   21-JUN-1999;  99US-0139817P.
PR   22-JUN-1999;  99US-0139899P.
PR   23-JUN-1999;  99US-0140353P.
PR   23-JUN-1999;  99US-0140354P.
PR   24-JUN-1999;  99US-0140695P.
PR   28-JUN-1999;  99US-0140823P.
PR   29-JUN-1999;  99US-0140991P.
PR   30-JUN-1999;  99US-0141287P.
PR   01-JUL-1999;  99US-0141842P.
PR   01-JUL-1999;  99US-0142154P.
PR   02-JUL-1999;  99US-0142055P.
PR   06-JUL-1999;  99US-0142390P.
PR   08-JUL-1999;  99US-0148803P.
PR   09-JUL-1999;  99US-0142920P.
PR   12-JUL-1999;  99US-0142977P.
PR   13-JUL-1999;  99US-0143542P.
PR   14-JUL-1999;  99US-0143624P.
PR   15-JUL-1999;  99US-0144005P.
PR   16-JUL-1999;  99US-0144085P.
PR   16-JUL-1999;  99US-0144086P.
PR   19-JUL-1999;  99US-0144325P.
PR   19-JUL-1999;  99US-0144331P.
PR   19-JUL-1999;  99US-0144332P.
PR   19-JUL-1999;  99US-0144333P.
PR   19-JUL-1999;  99US-0144334P.
PR   19-JUL-1999;  99US-0144335P.
PR   20-JUL-1999;  99US-0144352P.
PR   20-JUL-1999;  99US-0144632P.
PR   20-JUL-1999;  99US-0144484P.
PR   21-JUL-1999;  99US-0144814P.
PR   21-JUL-1999;  99US-0145086P.
PR   21-JUL-1999;  99US-0145088P.
PR   22-JUL-1999;  99US-0145085P.
PR   22-JUL-1999;  99US-0145087P.
PR   22-JUL-1999;  99US-0145089P.
PR   22-JUL-1999;  99US-0145192P.
PR   23-JUL-1999;  99US-0145145P.
PR   23-JUL-1999;  99US-0145218P.
PR   23-JUL-1999;  99US-0145224P.
PR   26-JUL-1999;  99US-0145276P.
PR   27-JUL-1999;  99US-0145913P.
PR   27-JUL-1999;  99US-0145918P.
PR   27-JUL-1999;  99US-0145919P.
PR   28-JUL-1999;  99US-0145951P.
PR   02-AUG-1999;  99US-0146386P.
PR   02-AUG-1999;  99US-0146388P.
PR   03-AUG-1999;  99US-0146389P.
PR   03-AUG-1999;  99US-0147038P.
PR   04-AUG-1999;  99US-0147204P.
PR   04-AUG-1999;  99US-0147302P.
PR   05-AUG-1999;  99US-0147192P.
PR   05-AUG-1999;  99US-0147260P.
PR   06-AUG-1999;  99US-0147303P.
PR   06-AUG-1999;  99US-0147416P.
PR   09-AUG-1999;  99US-0147493P.
PR   09-AUG-1999;  99US-0147935P.
PR   10-AUG-1999;  99US-0148171P.
PR   11-AUG-1999;  99US-0148319P.
PR   12-AUG-1999;  99US-0148341P.
PR   13-AUG-1999;  99US-0148565P.
PR   13-AUG-1999;  99US-0148684P.
```

PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0143390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0151192P.
PR 23-JUL-1999; 99US-0151145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 25-AUG-1999; 99US-0150566P.
PR 25-AUG-1999; 99US-0150884P.
PR 26-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.

PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match   50.8%;   Score 135; DB 3; Length 453;
Best Local Similarity   46.9%;   Pred. No. 1.2e-06;
Matches   23; Conservative   5; Mismatches 15; Indels   6; Gaps   1;

Qy   2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSS------LCPTCRCPV 44
        | ||::|| || | |:||||||||:|:||| |||||| |||||| |
Db 125 CSICLNIWHDVVTAAPCLHNFCNGCFSEWMRSEEKHKHVLCPQCRTTV 173

RESULT 15
AAG37658
ID AAG37658 standard; protein; 453 AA.
XX
AC AAG37658;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46342.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.

PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.

PN		EP1033405-A2.
XX		
PD		06-SEP-2000.
XX		
PF		25-FEB-2000; 2000EP-00301439.
XX		
PR		25-FEB-1999; 99US-0121825P.
PR		05-MAR-1999; 99US-0123180P.
PR		09-MAR-1999; 99US-0123548P.
PR		23-MAR-1999; 99US-0125788P.
PR		25-MAR-1999; 99US-0126264P.
PR		29-MAR-1999; 99US-0126785P.
PR		01-APR-1999; 99US-0127462P.
PR		06-APR-1999; 99US-0128234P.
PR		08-APR-1999; 99US-0128714P.
PR		16-APR-1999; 99US-0129845P.
PR		19-APR-1999; 99US-0130077P.
PR		21-APR-1999; 99US-0130449P.
PR		23-APR-1999; 99US-0130510P.
PR		23-APR-1999; 99US-0130891P.
PR		28-APR-1999; 99US-0131449P.
PR		30-APR-1999; 99US-0132048P.
PR		30-APR-1999; 99US-0132407P.
PR		04-MAY-1999; 99US-0132484P.
PR		05-MAY-1999; 99US-0132485P.
PR		06-MAY-1999; 99US-0132486P.
PR		06-MAY-1999; 99US-0132487P.
PR		07-MAY-1999; 99US-0132863P.
PR		11-MAY-1999; 99US-0134255P.
PR		14-MAY-1999; 99US-0134218P.
PR		14-MAY-1999; 99US-0134219P.
PR		14-MAY-1999; 99US-0134221P.
PR		14-MAY-1999; 99US-0134370P.
PR		18-MAY-1999; 99US-0134768P.
PR		19-MAY-1999; 99US-0134941P.
PR		20-MAY-1999; 99US-0135124P.
PR		21-MAY-1999; 99US-0135353P.
PR		24-MAY-1999; 99US-0135629P.
PR		25-MAY-1999; 99US-0136021P.
PR		27-MAY-1999; 99US-0136392P.
PR		28-MAY-1999; 99US-0136782P.
PR		01-JUN-1999; 99US-0137222P.
PR		03-JUN-1999; 99US-0137528P.
PR		04-JUN-1999; 99US-0137502P.
PR		07-JUN-1999; 99US-0137724P.
PR		08-JUN-1999; 99US-0138094P.
PR		10-JUN-1999; 99US-0138540P.
PR		10-JUN-1999; 99US-0138477P.
PR		14-JUN-1999; 99US-0139119P.
PR		16-JUN-1999; 99US-0139452P.
PR		16-JUN-1999; 99US-0139453P.
PR		17-JUN-1999; 99US-0139492P.
PR		18-JUN-1999; 99US-0139454P.
PR		18-JUN-1999; 99US-0139455P.
PR		18-JUN-1999; 99US-0139456P.
PR		18-JUN-1999; 99US-0139457P.
PR		18-JUN-1999; 99US-0139458P.
PR		18-JUN-1999; 99US-0139459P.
PR		18-JUN-1999; 99US-0139460P.
PR		18-JUN-1999; 99US-0139461P.
PR		18-JUN-1999; 99US-0139462P.
PR		18-JUN-1999; 99US-0139463P.
PR		18-JUN-1999; 99US-0139750P.
PR		18-JUN-1999; 99US-0139763P.
PR		21-JUN-1999; 99US-0139817P.
PR		22-JUN-1999; 99US-0139899P.
PR		23-JUN-1999; 99US-0140353P.
PR		23-JUN-1999; 99US-0140354P.
PR		24-JUN-1999; 99US-0140695P.
PR		28-JUN-1999; 99US-0140823P.
PR		29-JUN-1999; 99US-0140991P.
PR		30-JUN-1999; 99US-0141287P.
PR		01-JUL-1999; 99US-0141842P.

```
Query Match     50.8%;  Score 135;  DB 3;  Length 426;
Best Local Similarity  46.9%;  Pred. No. 1.2e-06;
Matches  23;  Conservative   5;  Mismatches  15;  Indels  6;  Gaps  1;

Qy     2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSS------LCPTCRCPV 44
         | ||| :|| ||:  |:|: ||:|  |:|:| |:|||    ||| |: ||
Db    98 CSICLNIWHDVVTAAPCLHNFCNGCPSEWMRRSEEKHKHVLCPQCRTTV 146
```

RESULT 14
AAG20352
ID AAG20352 standard; protein; 453 AA.
XX
AC AAG20352;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 22505.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX

```
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999; 99US-0121825P.          PR   23-JUN-1999; 99US-0140353P.
PR   05-MAR-1999; 99US-0123180P.          PR   23-JUN-1999; 99US-0140354P.
PR   09-MAR-1999; 99US-0123548P.          PR   24-JUN-1999; 99US-0140695P.
PR   23-MAR-1999; 99US-0125788P.          PR   28-JUN-1999; 99US-0140823P.
PR   25-MAR-1999; 99US-0126264P.          PR   29-JUN-1999; 99US-0140991P.
PR   29-MAR-1999; 99US-0126785P.          PR   30-JUN-1999; 99US-0141287P.
PR   01-APR-1999; 99US-0127462P.          PR   01-JUL-1999; 99US-0141842P.
PR   06-APR-1999; 99US-0128234P.          PR   01-JUL-1999; 99US-0142154P.
PR   08-APR-1999; 99US-0128714P.          PR   02-JUL-1999; 99US-0142055P.
PR   16-APR-1999; 99US-0129845P.          PR   06-JUL-1999; 99US-0142390P.
PR   19-APR-1999; 99US-0130077P.          PR   08-JUL-1999; 99US-0142803P.
PR   21-APR-1999; 99US-0130449P.          PR   09-JUL-1999; 99US-0142920P.
PR   23-APR-1999; 99US-0130510P.          PR   12-JUL-1999; 99US-0142977P.
PR   23-APR-1999; 99US-0130891P.          PR   13-JUL-1999; 99US-0143542P.
PR   28-APR-1999; 99US-0131449P.          PR   14-JUL-1999; 99US-0143624P.
PR   30-APR-1999; 99US-0132048P.          PR   15-JUL-1999; 99US-0144005P.
PR   30-APR-1999; 99US-0132407P.          PR   16-JUL-1999; 99US-0144085P.
PR   04-MAY-1999; 99US-0132484P.          PR   16-JUL-1999; 99US-0144086P.
PR   05-MAY-1999; 99US-0132485P.          PR   19-JUL-1999; 99US-0144325P.
PR   06-MAY-1999; 99US-0132486P.          PR   19-JUL-1999; 99US-0144331P.
PR   06-MAY-1999; 99US-0132487P.          PR   19-JUL-1999; 99US-0144332P.
PR   07-MAY-1999; 99US-0132863P.          PR   19-JUL-1999; 99US-0144333P.
PR   11-MAY-1999; 99US-0134256P.          PR   19-JUL-1999; 99US-0144334P.
PR   14-MAY-1999; 99US-0134218P.          PR   19-JUL-1999; 99US-0144335P.
PR   14-MAY-1999; 99US-0134219P.          PR   20-JUL-1999; 99US-0144352P.
PR   14-MAY-1999; 99US-0134221P.          PR   20-JUL-1999; 99US-0144632P.
PR   14-MAY-1999; 99US-0134370P.          PR   20-JUL-1999; 99US-0144884P.
PR   18-MAY-1999; 99US-0134768P.          PR   21-JUL-1999; 99US-0144814P.
PR   19-MAY-1999; 99US-0134941P.          PR   21-JUL-1999; 99US-0145086P.
PR   20-MAY-1999; 99US-0135124P.          PR   21-JUL-1999; 99US-0145088P.
PR   21-MAY-1999; 99US-0135353P.          PR   22-JUL-1999; 99US-0145085P.
PR   24-MAY-1999; 99US-0135629P.          PR   22-JUL-1999; 99US-0145087P.
PR   25-MAY-1999; 99US-0136021P.          PR   22-JUL-1999; 99US-0145089P.
PR   27-MAY-1999; 99US-0136392P.          PR   22-JUL-1999; 99US-0145192P.
PR   28-MAY-1999; 99US-0137282P.          PR   23-JUL-1999; 99US-0145145P.
PR   01-JUN-1999; 99US-0137222P.          PR   23-JUL-1999; 99US-0145218P.
PR   03-JUN-1999; 99US-0137528P.          PR   23-JUL-1999; 99US-0145224P.
PR   04-JUN-1999; 99US-0137502P.          PR   26-JUL-1999; 99US-0145276P.
PR   07-JUN-1999; 99US-0137724P.          PR   27-JUL-1999; 99US-0145913P.
PR   08-JUN-1999; 99US-0138094P.          PR   27-JUL-1999; 99US-0145918P.
PR   10-JUN-1999; 99US-0138540P.          PR   27-JUL-1999; 99US-0145919P.
PR   14-JUN-1999; 99US-0138847P.          PR   28-JUL-1999; 99US-0145951P.
PR   14-JUN-1999; 99US-0139119P.          PR   02-AUG-1999; 99US-0146386P.
PR   16-JUN-1999; 99US-0139452P.          PR   02-AUG-1999; 99US-0146388P.
PR   17-JUN-1999; 99US-0139453P.          PR   02-AUG-1999; 99US-0146389P.
PR   18-JUN-1999; 99US-0139492P.          PR   03-AUG-1999; 99US-0147038P.
PR   18-JUN-1999; 99US-0139454P.          PR   04-AUG-1999; 99US-0147204P.
PR   18-JUN-1999; 99US-0139455P.          PR   04-AUG-1999; 99US-0147302P.
PR   18-JUN-1999; 99US-0139456P.          PR   05-AUG-1999; 99US-0147192P.
PR   18-JUN-1999; 99US-0139457P.          PR   05-AUG-1999; 99US-0147260P.
PR   18-JUN-1999; 99US-0139458P.          PR   06-AUG-1999; 99US-0147303P.
PR   18-JUN-1999; 99US-0139459P.          PR   06-AUG-1999; 99US-0147416P.
PR   18-JUN-1999; 99US-0139460P.          PR   09-AUG-1999; 99US-0147493P.
PR   18-JUN-1999; 99US-0139461P.          PR   09-AUG-1999; 99US-0147935P.
PR   18-JUN-1999; 99US-0139462P.          PR   10-AUG-1999; 99US-0148171P.
PR   18-JUN-1999; 99US-0139463P.          PR   11-AUG-1999; 99US-0148319P.
PR   18-JUN-1999; 99US-0139750P.          PR   12-AUG-1999; 99US-0148341P.
PR   18-JUN-1999; 99US-0139763P.          PR   13-AUG-1999; 99US-0148565P.
PR   21-JUN-1999; 99US-0139817P.          PR   13-AUG-1999; 99US-0148684P.
PR   22-JUN-1999; 99US-0139899P.          PR   16-AUG-1999; 99US-0149368P.
                                          PR   17-AUG-1999; 99US-0149175P.
                                          PR   18-AUG-1999; 99US-0149426P.
                                          PR   20-AUG-1999; 99US-0149722P.
                                          PR   20-AUG-1999; 99US-0149723P.
                                          PR   20-AUG-1999; 99US-0149929P.
                                          PR   23-AUG-1999; 99US-0149902P.
                                          PR   23-AUG-1999; 99US-0149930P.
                                          PR   25-AUG-1999; 99US-0150566P.
                                          PR   26-AUG-1999; 99US-0150884P.
                                          PR   27-AUG-1999; 99US-0151065P.
                                          PR   27-AUG-1999; 99US-0151066P.
                                          PR   27-AUG-1999; 99US-0151080P.
                                          PR   30-AUG-1999; 99US-0151303P.
```

```
PR  18-JUN-1999;  99US-0139461P.
PR  18-JUN-1999;  99US-0139462P.
PR  18-JUN-1999;  99US-0139463P.
PR  18-JUN-1999;  99US-0139750P.
PR  18-JUN-1999;  99US-0139763P.
PR  21-JUN-1999;  99US-0139817P.
PR  22-JUN-1999;  99US-0139899P.
PR  23-JUN-1999;  99US-0140353P.
PR  23-JUN-1999;  99US-0140354P.
PR  24-JUN-1999;  99US-0140695P.
PR  28-JUN-1999;  99US-0140823P.
PR  29-JUN-1999;  99US-0140991P.
PR  30-JUN-1999;  99US-0141287P.
PR  01-JUL-1999;  99US-0141842P.
PR  01-JUL-1999;  99US-0142154P.
PR  02-JUL-1999;  99US-0142055P.
PR  06-JUL-1999;  99US-0142390P.
PR  08-JUL-1999;  99US-0142803P.
PR  09-JUL-1999;  99US-0142920P.
PR  12-JUL-1999;  99US-0142977P.
PR  13-JUL-1999;  99US-0143542P.
PR  14-JUL-1999;  99US-0143624P.
PR  15-JUL-1999;  99US-0144005P.
PR  16-JUL-1999;  99US-0144085P.
PR  16-JUL-1999;  99US-0144086P.
PR  19-JUL-1999;  99US-0144325P.
PR  19-JUL-1999;  99US-0144331P.
PR  19-JUL-1999;  99US-0144332P.
PR  19-JUL-1999;  99US-0144333P.
PR  19-JUL-1999;  99US-0144334P.
PR  20-JUL-1999;  99US-0144335P.
PR  20-JUL-1999;  99US-0144352P.
PR  20-JUL-1999;  99US-0144632P.
PR  20-JUL-1999;  99US-0144884P.
PR  21-JUL-1999;  99US-0144814P.
PR  21-JUL-1999;  99US-0145086P.
PR  21-JUL-1999;  99US-0145088P.
PR  22-JUL-1999;  99US-0145085P.
PR  22-JUL-1999;  99US-0145087P.
PR  22-JUL-1999;  99US-0145089P.
PR  22-JUL-1999;  99US-0145192P.
PR  23-JUL-1999;  99US-0145145P.
PR  23-JUL-1999;  99US-0145218P.
PR  23-JUL-1999;  99US-0145224P.
PR  26-JUL-1999;  99US-0145276P.
PR  27-JUL-1999;  99US-0145913P.
PR  27-JUL-1999;  99US-0145918P.
PR  27-JUL-1999;  99US-0145919P.
PR  28-JUL-1999;  99US-0145951P.
PR  02-AUG-1999;  99US-0146386P.
PR  02-AUG-1999;  99US-0146388P.
PR  02-AUG-1999;  99US-0146389P.
PR  03-AUG-1999;  99US-0147038P.
PR  04-AUG-1999;  99US-0147204P.
PR  04-AUG-1999;  99US-0147302P.
PR  05-AUG-1999;  99US-0147192P.
PR  05-AUG-1999;  99US-0147260P.
PR  06-AUG-1999;  99US-0147303P.
PR  06-AUG-1999;  99US-0147416P.
PR  09-AUG-1999;  99US-0147493P.
PR  09-AUG-1999;  99US-0147935P.
PR  10-AUG-1999;  99US-0148171P.
PR  11-AUG-1999;  99US-0148319P.
PR  12-AUG-1999;  99US-0148341P.
PR  13-AUG-1999;  99US-0148565P.
PR  13-AUG-1999;  99US-0148684P.
PR  16-AUG-1999;  99US-0149368P.
PR  17-AUG-1999;  99US-0149175P.
PR  18-AUG-1999;  99US-0149426P.
PR  20-AUG-1999;  99US-0149722P.
PR  20-AUG-1999;  99US-0149723P.
PR  20-AUG-1999;  99US-0149929P.
PR  20-AUG-1999;  99US-0149902P.

PR  23-AUG-1999;  99US-0149930P.
PR  25-AUG-1999;  99US-0150566P.
PR  26-AUG-1999;  99US-0150884P.
PR  27-AUG-1999;  99US-0151065P.
PR  27-AUG-1999;  99US-0151066P.
PR  27-AUG-1999;  99US-0151080P.
PR  30-AUG-1999;  99US-0151303P.
PR  31-AUG-1999;  99US-0151438P.
PR  01-SEP-1999;  99US-0151930P.
PR  07-SEP-1999;  99US-0152363P.
PR  10-SEP-1999;  99US-0153070P.
PR  13-SEP-1999;  99US-0153758P.
PR  15-SEP-1999;  99US-0154018P.
PR  16-SEP-1999;  99US-0154039P.
PR  20-SEP-1999;  99US-0154779P.
PR  22-SEP-1999;  99US-0155139P.
PR  23-SEP-1999;  99US-0154486P.
PR  24-SEP-1999;  99US-0155659P.
PR  28-SEP-1999;  99US-0156458P.
PR  29-SEP-1999;  99US-0156596P.
PR  04-OCT-1999;  99US-0157117P.
PR  05-OCT-1999;  99US-0157753P.
PR  06-OCT-1999;  99US-0157865P.
PR  07-OCT-1999;  99US-0158029P.
PR  08-OCT-1999;  99US-0158232P.
PR  12-OCT-1999;  99US-0158369P.
PR  13-OCT-1999;  99US-0159293P.
PR  13-OCT-1999;  99US-0159294P.
PR  13-OCT-1999;  99US-0159295P.
PR  14-OCT-1999;  99US-0159329P.
PR  14-OCT-1999;  99US-0159330P.
PR  14-OCT-1999;  99US-0159331P.
PR  14-OCT-1999;  99US-0159637P.
PR  14-OCT-1999;  99US-0159638P.
PR  18-OCT-1999;  99US-0159584P.
PR  21-OCT-1999;  99US-0160741P.
PR  21-OCT-1999;  99US-0160767P.
PR  21-OCT-1999;  99US-0160768P.
PR  21-OCT-1999;  99US-0160770P.
PR  21-OCT-1999;  99US-0160814P.
PR  21-OCT-1999;  99US-0160815P.
PR  22-OCT-1999;  99US-0160980P.
PR  22-OCT-1999;  99US-0160981P.
PR  22-OCT-1999;  99US-0160989P.
PR  25-OCT-1999;  99US-0161404P.
PR  25-OCT-1999;  99US-0161405P.
PR  25-OCT-1999;  99US-0161406P.
PR  26-OCT-1999;  99US-0161359P.
PR  26-OCT-1999;  99US-0161360P.
PR  26-OCT-1999;  99US-0161361P.
PR  28-OCT-1999;  99US-0161920P.
PR  28-OCT-1999;  99US-0161992P.
PR  28-OCT-1999;  99US-0161993P.
PR  29-OCT-1999;  99US-0162142P.
```

```
Query Match             50.8%;  Score 135;  DB 3;  Length 426;
Best Local Similarity   46.9%;  Pred. No. 1.2e-06;
Matches   23;  Conservative   5;  Mismatches  15;  Indels  6;  Gaps  1;

Qy          2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSS------LCPTCRCPV  44
              | |: |:: :|| ||  |:|  | |::|| || ||        ||| ||
Db         98 CSICLNIWHDVVTAAPCLHNFCNGCFSEWMRRSEEKHKGHVLCPQCRTTV 146


RESULT 13
AAG37659
ID   AAG37659 standard; protein; 426 AA.
XX
AC   AAG37659;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 46343.
```

```
ID   AAG20353 standard; protein; 426 AA.
XX
AC   AAG20353;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 22506.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-00301439.
XX
PR   25-FEB-1999; 99US-0121825P.
PR   05-MAR-1999; 99US-0123180P.
PR   09-MAR-1999; 99US-0123548P.
PR   23-MAR-1999; 99US-0125788P.
PR   25-MAR-1999; 99US-0126264P.
PR   29-MAR-1999; 99US-0126785P.
PR   01-APR-1999; 99US-0127462P.
PR   06-APR-1999; 99US-0128234P.
PR   08-APR-1999; 99US-0128714P.
PR   16-APR-1999; 99US-0129845P.
PR   19-APR-1999; 99US-0130077P.
PR   21-APR-1999; 99US-0130449P.
PR   23-APR-1999; 99US-0130510P.
PR   23-APR-1999; 99US-0130891P.
PR   28-APR-1999; 99US-0131449P.
PR   30-APR-1999; 99US-0132048P.
PR   30-APR-1999; 99US-0132407P.
PR   04-MAY-1999; 99US-0132484P.
PR   05-MAY-1999; 99US-0132485P.
PR   06-MAY-1999; 99US-0132486P.
PR   06-MAY-1999; 99US-0132487P.
PR   07-MAY-1999; 99US-0132863P.
PR   11-MAY-1999; 99US-0134256P.
PR   14-MAY-1999; 99US-0134218P.
PR   14-MAY-1999; 99US-0134219P.
PR   14-MAY-1999; 99US-0134221P.
PR   14-MAY-1999; 99US-0134370P.
PR   18-MAY-1999; 99US-0134768P.
PR   19-MAY-1999; 99US-0134941P.
PR   20-MAY-1999; 99US-0135124P.
PR   21-MAY-1999; 99US-0135353P.
PR   24-MAY-1999; 99US-0135629P.
PR   25-MAY-1999; 99US-0136021P.
PR   27-MAY-1999; 99US-0136392P.
PR   28-MAY-1999; 99US-0136782P.
PR   01-JUN-1999; 99US-0137222P.
PR   03-JUN-1999; 99US-0137528P.
PR   04-JUN-1999; 99US-0137502P.
PR   07-JUN-1999; 99US-0137724P.
PR   08-JUN-1999; 99US-0138094P.
PR   10-JUN-1999; 99US-0138540P.
PR   10-JUN-1999; 99US-0138847P.
PR   14-JUN-1999; 99US-0139119P.
PR   16-JUN-1999; 99US-0139452P.
PR   16-JUN-1999; 99US-0139453P.
PR   17-JUN-1999; 99US-0139492P.
PR   18-JUN-1999; 99US-0139454P.
PR   18-JUN-1999; 99US-0139455P.
PR   18-JUN-1999; 99US-0139456P.
PR   18-JUN-1999; 99US-0139457P.
PR   18-JUN-1999; 99US-0139458P.
PR   18-JUN-1999; 99US-0139459P.
PR   18-JUN-1999; 99US-0139460P.
PR   16-AUG-1999; 99US-0149368P.
PR   17-AUG-1999; 99US-0149175P.
PR   18-AUG-1999; 99US-0149426P.
PR   20-AUG-1999; 99US-0149722P.
PR   20-AUG-1999; 99US-0149723P.
PR   20-AUG-1999; 99US-0149929P.
PR   23-AUG-1999; 99US-0149902P.
PR   23-AUG-1999; 99US-0149930P.
PR   25-AUG-1999; 99US-0150566P.
PR   26-AUG-1999; 99US-0150884P.
PR   27-AUG-1999; 99US-0151065P.
PR   27-AUG-1999; 99US-0151066P.
PR   27-AUG-1999; 99US-0151080P.
PR   30-AUG-1999; 99US-0151303P.
PR   31-AUG-1999; 99US-0151438P.
PR   01-SEP-1999; 99US-0151930P.
PR   07-SEP-1999; 99US-0152363P.
PR   10-SEP-1999; 99US-0153070P.
PR   13-SEP-1999; 99US-0153758P.
PR   15-SEP-1999; 99US-0154018P.
PR   16-SEP-1999; 99US-0154039P.
PR   20-SEP-1999; 99US-0154779P.
PR   22-SEP-1999; 99US-0155139P.
PR   23-SEP-1999; 99US-0155486P.
PR   24-SEP-1999; 99US-0155659P.
PR   28-SEP-1999; 99US-0156458P.
PR   29-SEP-1999; 99US-0156596P.
PR   04-OCT-1999; 99US-0157117P.
PR   05-OCT-1999; 99US-0157753P.
PR   06-OCT-1999; 99US-0157865P.
PR   07-OCT-1999; 99US-0158029P.
PR   08-OCT-1999; 99US-0158232P.
PR   12-OCT-1999; 99US-0158369P.
PR   13-OCT-1999; 99US-0159293P.
PR   13-OCT-1999; 99US-0159294P.
PR   13-OCT-1999; 99US-0159295P.
PR   13-OCT-1999; 99US-0159329P.
PR   14-OCT-1999; 99US-0159330P.
PR   14-OCT-1999; 99US-0159331P.
PR   14-OCT-1999; 99US-0159637P.
PR   14-OCT-1999; 99US-0159638P.
PR   14-OCT-1999; 99US-0159584P.
PR   18-OCT-1999; 99US-0160741P.
PR   21-OCT-1999; 99US-0160767P.
PR   21-OCT-1999; 99US-0160768P.
PR   21-OCT-1999; 99US-0160770P.
PR   21-OCT-1999; 99US-0160814P.
PR   21-OCT-1999; 99US-0160815P.
PR   22-OCT-1999; 99US-0160980P.
PR   22-OCT-1999; 99US-0160981P.
PR   22-OCT-1999; 99US-0160989P.
PR   25-OCT-1999; 99US-0161404P.
PR   25-OCT-1999; 99US-0161405P.
PR   25-OCT-1999; 99US-0161406P.
PR   26-OCT-1999; 99US-0161359P.
PR   26-OCT-1999; 99US-0161360P.
PR   26-OCT-1999; 99US-0161361P.
PR   28-OCT-1999; 99US-0161920P.
PR   28-OCT-1999; 99US-0161932P.
PR   28-OCT-1999; 99US-0161993P.
PR   29-OCT-1999; 99US-0162142P.
```

```
Query Match             50.8%;  Score 135;  DB 3;  Length 350;
Best Local Similarity   46.9%;  Pred. No. 9.9e-07;
Matches   23;  Conservative   5;  Mismatches  15;  Indels   6;  Gaps   1;

Qy    2 CIICQDLLHDCVSLQPCMHTFCAACYSGWMERSS------LCPTCRCPV 44
          |||| || |||: :||| | :|:||: |||:||      |||| || |
Db   22 CSICLNIWHDVVTAAPCLRNFCNGCFSEWMRRSEEKHKHVLCPQCRRTTV 70
```

RESULT 12
AAG20353

PR	08-JUN-1999;	99US-0138094P.	PR	09-AUG-1999;	99US-0147493P.
PR	10-JUN-1999;	99US-0138540P.	PR	09-AUG-1999;	99US-0147935P.
PR	10-JUN-1999;	99US-0138847P.	PR	10-AUG-1999;	99US-0148171P.
PR	14-JUN-1999;	99US-0139119P.	PR	11-AUG-1999;	99US-0148319P.
PR	16-JUN-1999;	99US-0139452P.	PR	12-AUG-1999;	99US-0148341P.
PR	16-JUN-1999;	99US-0139453P.	PR	13-AUG-1999;	99US-0148565P.
PR	17-JUN-1999;	99US-0139492P.	PR	13-AUG-1999;	99US-0148684P.
PR	18-JUN-1999;	99US-0139454P.	PR	16-AUG-1999;	99US-0149368P.
PR	18-JUN-1999;	99US-0139455P.	PR	17-AUG-1999;	99US-0149175P.
PR	18-JUN-1999;	99US-0139456P.	PR	18-AUG-1999;	99US-0149426P.
PR	18-JUN-1999;	99US-0139457P.	PR	20-AUG-1999;	99US-0149722P.
PR	18-JUN-1999;	99US-0139458P.	PR	20-AUG-1999;	99US-0149723P.
PR	18-JUN-1999;	99US-0139459P.	PR	20-AUG-1999;	99US-0149929P.
PR	18-JUN-1999;	99US-0139460P.	PR	23-AUG-1999;	99US-0149902P.
PR	18-JUN-1999;	99US-0139461P.	PR	23-AUG-1999;	99US-0149930P.
PR	18-JUN-1999;	99US-0139462P.	PR	25-AUG-1999;	99US-0150566P.
PR	18-JUN-1999;	99US-0139463P.	PR	26-AUG-1999;	99US-0150884P.
PR	18-JUN-1999;	99US-0139763P.	PR	27-AUG-1999;	99US-0151065P.
PR	21-JUN-1999;	99US-0139817P.	PR	27-AUG-1999;	99US-0151066P.
PR	22-JUN-1999;	99US-0139899P.	PR	30-AUG-1999;	99US-0151080P.
PR	23-JUN-1999;	99US-0140353P.	PR	31-AUG-1999;	99US-0151303P.
PR	23-JUN-1999;	99US-0140354P.	PR	01-SEP-1999;	99US-0151930P.
PR	24-JUN-1999;	99US-0140695P.	PR	07-SEP-1999;	99US-0151438P.
PR	28-JUN-1999;	99US-0140823P.	PR	10-SEP-1999;	99US-0152263P.
PR	29-JUN-1999;	99US-0140991P.	PR	13-SEP-1999;	99US-0153070P.
PR	30-JUN-1999;	99US-0141287P.	PR	13-SEP-1999;	99US-0153758P.
PR	01-JUL-1999;	99US-0141842P.	PR	15-SEP-1999;	99US-0154018P.
PR	01-JUL-1999;	99US-0142154P.	PR	16-SEP-1999;	99US-0154039P.
PR	02-JUL-1999;	99US-0142055P.	PR	20-SEP-1999;	99US-0154779P.
PR	06-JUL-1999;	99US-0142390P.	PR	22-SEP-1999;	99US-0155139P.
PR	08-JUL-1999;	99US-0142803P.	PR	23-SEP-1999;	99US-0155486P.
PR	09-JUL-1999;	99US-0142920P.	PR	24-SEP-1999;	99US-0155659P.
PR	12-JUL-1999;	99US-0142977P.	PR	28-SEP-1999;	99US-0154458P.
PR	13-JUL-1999;	99US-0143542P.	PR	29-SEP-1999;	99US-0155596P.
PR	14-JUL-1999;	99US-0143624P.	PR	04-OCT-1999;	99US-0157117P.
PR	15-JUL-1999;	99US-0144005P.	PR	05-OCT-1999;	99US-0157753P.
PR	16-JUL-1999;	99US-0144085P.	PR	06-OCT-1999;	99US-0157865P.
PR	16-JUL-1999;	99US-0144086P.	PR	07-OCT-1999;	99US-0158029P.
PR	19-JUL-1999;	99US-0144325P.	PR	08-OCT-1999;	99US-0158232P.
PR	19-JUL-1999;	99US-0144331P.	PR	12-OCT-1999;	99US-0158369P.
PR	19-JUL-1999;	99US-0144332P.	PR	13-OCT-1999;	99US-0159293P.
PR	19-JUL-1999;	99US-0144333P.	PR	13-OCT-1999;	99US-0159294P.
PR	19-JUL-1999;	99US-0144334P.	PR	13-OCT-1999;	99US-0159295P.
PR	19-JUL-1999;	99US-0144335P.	PR	14-OCT-1999;	99US-0159329P.
PR	20-JUL-1999;	99US-0144352P.	PR	14-OCT-1999;	99US-0159330P.
PR	20-JUL-1999;	99US-0144632P.	PR	14-OCT-1999;	99US-0159331P.
PR	20-JUL-1999;	99US-0144884P.	PR	14-OCT-1999;	99US-0159637P.
PR	21-JUL-1999;	99US-0144814P.	PR	14-OCT-1999;	99US-0159638P.
PR	21-JUL-1999;	99US-0145086P.	PR	18-OCT-1999;	99US-0159584P.
PR	21-JUL-1999;	99US-0145088P.	PR	21-OCT-1999;	99US-0160741P.
PR	22-JUL-1999;	99US-0145085P.	PR	21-OCT-1999;	99US-0160767P.
PR	22-JUL-1999;	99US-0145087P.	PR	21-OCT-1999;	99US-0160768P.
PR	22-JUL-1999;	99US-0145089P.	PR	21-OCT-1999;	99US-0160770P.
PR	22-JUL-1999;	99US-0145192P.	PR	21-OCT-1999;	99US-0160814P.
PR	23-JUL-1999;	99US-0145145P.	PR	21-OCT-1999;	99US-0160815P.
PR	23-JUL-1999;	99US-0145218P.	PR	22-OCT-1999;	99US-0160800P.
PR	23-JUL-1999;	99US-0145224P.	PR	22-OCT-1999;	99US-0160981P.
PR	26-JUL-1999;	99US-0152276P.	PR	22-OCT-1999;	99US-0160989P.
PR	27-JUL-1999;	99US-0145913P.	PR	25-OCT-1999;	99US-0161404P.
PR	27-JUL-1999;	99US-0145918P.	PR	25-OCT-1999;	99US-0161405P.
PR	27-JUL-1999;	99US-0145919P.	PR	25-OCT-1999;	99US-0161406P.
PR	28-JUL-1999;	99US-0145951P.	PR	26-OCT-1999;	99US-0161359P.
PR	02-AUG-1999;	99US-0146386P.	PR	26-OCT-1999;	99US-0161360P.
PR	02-AUG-1999;	99US-0146388P.	PR	26-OCT-1999;	99US-0161361P.
PR	02-AUG-1999;	99US-0146389P.	PR	28-OCT-1999;	99US-0161920P.
PR	03-AUG-1999;	99US-0147038P.	PR	28-OCT-1999;	99US-0161992P.
PR	04-AUG-1999;	99US-0147204P.	PR	28-OCT-1999;	99US-0161993P.
PR	04-AUG-1999;	99US-0147302P.	PR	29-OCT-1999;	99US-0162142P.
PR	05-AUG-1999;	99US-0147192P.			
PR	05-AUG-1999;	99US-0147260P.			
PR	06-AUG-1999;	99US-0147303P.			
PR	06-AUG-1999;	99US-0147416P.			

Query Match   50.8%; Score 135; DB 3; Length 350;
Best Local Similarity   46.9%; Pred. No. 9.9e-07;
Matches   23; Conservative   5; Mismatches 15; Indels   6; Gaps   1;

```
PR    26-JUL-2000;    2000US-0220964P.
PR    14-AUG-2000;    2000US-0224518P.
PR    14-AUG-2000;    2000US-0224519P.
PR    14-AUG-2000;    2000US-0225267P.
PR    14-AUG-2000;    2000US-0225268P.
PR    14-AUG-2000;    2000US-0225270P.
PR    14-AUG-2000;    2000US-0225447P.
PR    14-AUG-2000;    2000US-0225757P.
PR    14-AUG-2000;    2000US-0225758P.
PR    22-AUG-2000;    2000US-0226868P.
PR    30-AUG-2000;    2000US-0228924P.
PR    01-SEP-2000;    2000US-0229287P.
PR    01-SEP-2000;    2000US-0229343P.
PR    01-SEP-2000;    2000US-0229344P.
PR    01-SEP-2000;    2000US-0229345P.
PR    05-SEP-2000;    2000US-0229509P.
PR    05-SEP-2000;    2000US-0229513P.
PR    08-SEP-2000;    2000US-0231413P.
PR    21-SEP-2000;    2000US-0234223P.
PR    21-SEP-2000;    2000US-0234274P.
PR    25-SEP-2000;    2000US-0234997P.
PR    27-SEP-2000;    2000US-0235834P.
PR    29-SEP-2000;    2000US-0236327P.
PR    29-SEP-2000;    2000US-0236367P.
PR    29-SEP-2000;    2000US-0236368P.
PR    29-SEP-2000;    2000US-0236369P.
PR    29-SEP-2000;    2000US-0236370P.
PR    02-OCT-2000;    2000US-0236802P.
PR    02-OCT-2000;    2000US-0237037P.
PR    02-OCT-2000;    2000US-0237038P.
PR    02-OCT-2000;    2000US-0237039P.
PR    02-OCT-2000;    2000US-0237040P.
PR    13-OCT-2000;    2000US-0239935P.
PR    20-OCT-2000;    2000US-0240960P.
PR    20-OCT-2000;    2000US-0241785P.
PR    20-OCT-2000;    2000US-0241809P.
PR    01-NOV-2000;    2000US-0244617P.
PR    17-NOV-2000;    2000US-0249299P.
PR    08-DEC-2000;    2000US-0251156P.
PR    08-DEC-2000;    2000US-0251868P.
PR    08-DEC-2000;    2000US-0251869P.
XX
PA    (ROSE/)  ROSEN C A.
PA    (RUBE/)  RUBEN S M.
PA    (BARA/)  BARASH S C.
XX
PI    Rosen CA, Ruben SM, Barash SC;
XX
DR    WPI; 2003-147444/14.
DR    N-PSDB; ABX73645.
XX
PT    New polypeptides and nucleic acids, useful in gene therapy for treating,
PT    inhibiting or preventing e.g. neural, immune system, muscular,
PT    respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT    renal disorders.
XX
PS    Claim 11; SEQ ID NO 1270; 402pp; English.
XX
CC    The invention relates to human novel polypeptides and their associated
CC    polynucleotides. The polypeptides and polynucleotides are useful in gene
CC    therapy for treating, inhibiting or preventing neural disorders, immune
CC    system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC    and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC    nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC    gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC    (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC    heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC    renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC    leukaemia), inflammatory diseases and conditions (e.g. asthma), blood
CC    appendicitis), allergic reactions and conditions (e.g. septic shock, bursitis and
CC    related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC    infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC    ABU55748 represent human novel polypeptides of the invention
```

```
XX
SQ    Sequence 230 AA;

      Query Match             66.2%;  Score 176;  DB 6;  Length 230;
      Best Local Similarity  100.0%;  Pred. No. 2.7e-11;
      Matches   28;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       17  PCMHTFCAACYSGWMERSSLCPTCRCPV  44
             ||||||||||||||||||||||||||||
Db        5  PCMHTFCAACYSGWMERSSLCPTCRCPV  32

RESULT 10
AAG20354
ID    AAG20354 standard; protein; 350 AA.
XX
AC    AAG20354;
XX
DT    17-OCT-2000  (first entry)
XX
DE    Arabidopsis thaliana protein fragment SEQ ID NO: 22507.
XX
KW    Protein identification; signal transduction pathway; metabolic pathway;
KW    hybridisation assay; genetic mapping; gene expression control; promoter;
KW    termination sequence.
XX
OS    Arabidopsis thaliana.
XX
PN    EP1033405-A2.
XX
PD    06-SEP-2000.
XX
PF    25-FEB-2000;  2000EP-00301439.
XX
PR    25-FEB-1999;  99US-0121825P.
PR    05-MAR-1999;  99US-0123180P.
PR    09-MAR-1999;  99US-0123548P.
PR    23-MAR-1999;  99US-0125788P.
PR    25-MAR-1999;  99US-0126264P.
PR    29-MAR-1999;  99US-0126785P.
PR    01-APR-1999;  99US-0127462P.
PR    06-APR-1999;  99US-0128234P.
PR    08-APR-1999;  99US-0128714P.
PR    16-APR-1999;  99US-0129845P.
PR    19-APR-1999;  99US-0130077P.
PR    21-APR-1999;  99US-0130449P.
PR    23-APR-1999;  99US-0130510P.
PR    23-APR-1999;  99US-0130891P.
PR    28-APR-1999;  99US-0131449P.
PR    30-APR-1999;  99US-0132048P.
PR    30-APR-1999;  99US-0132407P.
PR    04-MAY-1999;  99US-0132484P.
PR    05-MAY-1999;  99US-0132485P.
PR    06-MAY-1999;  99US-0132486P.
PR    06-MAY-1999;  99US-0132487P.
PR    07-MAY-1999;  99US-0132863P.
PR    11-MAY-1999;  99US-0134256P.
PR    14-MAY-1999;  99US-0134218P.
PR    14-MAY-1999;  99US-0134219P.
PR    14-MAY-1999;  99US-0134221P.
PR    14-MAY-1999;  99US-0134370P.
PR    18-MAY-1999;  99US-0134768P.
PR    19-MAY-1999;  99US-0134941P.
PR    20-MAY-1999;  99US-0135124P.
PR    21-MAY-1999;  99US-0135353P.
PR    24-MAY-1999;  99US-0135629P.
PR    25-MAY-1999;  99US-0136021P.
PR    27-MAY-1999;  99US-0136392P.
PR    28-MAY-1999;  99US-0136782P.
PR    01-JUN-1999;  99US-0137222P.
PR    03-JUN-1999;  99US-0137528P.
PR    04-JUN-1999;  99US-0137529P.
PR    07-JUN-1999;  99US-0137724P.
```

PT   used as food additives or preservatives.
XX
PS   Claim 11; SEQ ID NO 1270; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed

Query Match   66.2%;   Score 176;   DB 4; Length 230;
Best Local Similarity   100.0%;   Pred. No. 2.7e-11;
Matches   28;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   17   PCMHTFCAACYSGWMERSSLCPTCRCPV 44
           ||||||||||||||||||||||||||||
Db   5   PCMHTFCAACYSGWMERSSLCPTCRCPV 32

RESULT 9
ABU55386
ID   ABU55386 standard; protein; 230 AA.
XX
AC   ABU55386;
XX
DT   18-MAR-2003   (first entry)
XX
DE   Human novel polypeptide #473.
XX
KW   Human; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0215880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.

--- (left column) ---

PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0246617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   N-PSDB; AAS26304.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and

DR   WPI; 2001-182927/18.
DR   N-PSDB; AAF30352.
XX
PT   Novel nucleic acid sequence of mitotic checkpoint gene encoding a
PT   checkpoint with forkhead-associated domain and ring finger protein, for
PT   diagnosing tumorigenic cells and in screening for anticancer drugs.
XX
PS   Claim 8(a); Fig 4A-C; 85pp; English.
XX
CC   The present sequence is that of human mitotic checkpoint protein Chfr,
CC   having a forkhead associated domain (FHA) and a ring finger domain. The
CC   protein is required for regulation of the transition of cells from
CC   prophase to metaphase during mitosis. It has ubiquitin-protein ligase
CC   activity. The Chfr checkpoint was evident in primary human cells, but was
CC   inactivated in 4 of 8 human cancer cell lines. In U2OS cells, a mutation
CC   was identified that caused a Val to Met amino acid substitution in the
CC   highly conserved C-terminal Cys-rich region of the Chfr protein. In the
CC   absence of the Chfr checkpoint, cells subjected to mitotic stress
CC   condensed their chromosomes despite failing to separate their
CC   chromosomes. Chfr may monitor centrosome separation. Inactivation of the
CC   chfr gene (see AAF30352) in human cancer is theorized to underlie the
CC   increased sensitivity of cancer cells to antimitotic drugs. Polypeptides
CC   comprising the present sequence, or sequences comprising at least amino
CC   acids 31-103, 303-346 and/or 476-641 of this sequence, are claimed.
CC   Claimed methods of determining the tumourigenic potential of a cell
CC   comprise examining the cell for the presence of Chfr expression or for
CC   the presence of Chfr-mediated ubiquitin-protein ligase activity (in both
CC   cases, absence of expression indicating predisposition to tumourigenesis
CC   upon exposure to mitotic stress). A diagnostic kit for detecting the
CC   tumourigenic potential of cell cells comprises may comprise a ligand that
CC   binds to Chfr, such as an antibody or its fragment. Inhibitors of Chfr
CC   are identified by monitoring their effect on Chfr expression, and are
CC   used to retard the growth of cancer cells
XX
SQ   Sequence 664 AA;

Query Match   100.0%; Score 266; DB 4; Length 664;
Best Local Similarity   100.0%; Pred. No. 1.3e-20;
Matches   44; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSLCPTCRCPV 44
          ||||||||||||||||||||||||||||||||||||||||||||
Db   303 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSLCPTCRCPV 346

RESULT 8
AAU16317
ID   AAU16317 standard; protein; 230 AA.
XX
AC   AAU16317;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human novel secreted protein, Seq ID 1270.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX

PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0185350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   07-JUN-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0231968P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231966P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234999P.
PR   25-SEP-2000; 2000US-0234999P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235836P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.

```
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000;   2000JP-00118776.
PR   02-MAY-2000;   2000JP-00183767.
PR   09-JUN-2000;   2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 12100; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 652 AA;
XX
Query Match        100.0%;  Score 266;  DB 4;  Length 652;
Best Local Similarity  100.0%;  Pred. No. 1.3e-20;
Matches  44;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||||||||||||||||||||||||||||||||||||||||||
Db  291 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 334

RESULT 6
AAB83843
ID   AAB83843 standard; protein; 664 AA.
XX
AC   AAB83843;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Amino acid sequence of a human ring finger protein designated FHAR1.
XX
KW   FHAR1; RING finger protein; cancer; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200142430-A1.
XX
PD   14-JUN-2001.
XX
PF   07-DEC-2000; 2000WO-US033094.
```

```
XX
PR   08-DEC-1999;   99US-00456876.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Zhou B, Zhu Y, Chaturvedi P, Hurle MR, Li X;
XX
DR   WPI; 2001-381663/40.
DR   N-PSDB; AAF89709.
XX
PT   New FHAR1 polypeptide, a member of the RING finger protein family for
PT   diagnosing and treating cancer, and for use in anti-cancer vaccines.
XX
PS   Claim 1; Page 19; 28pp; English.
XX
CC   The present sequence represents a FHAR1 polypeptide, which is a member of
CC   the RING finger protein family. FHAR1 is useful in the treatment of
CC   cancer, and as a vaccine for inducing an immunological response in a
CC   mammal. FHAR1 polynucleotides may also be used as a diagnostic reagent
CC   through detection of mutations in the associated gene, and for chromosome
CC   localization studies, and tissue expression studies. FHAR1 antibodies are
CC   useful to isolate and to identify clones expressing the polypeptides, or
CC   to purify the polypeptides by affinity chromatography and to treat cancer
XX
SQ   Sequence 664 AA;
XX
Query Match        100.0%;  Score 266;  DB 4;  Length 664;
Best Local Similarity  100.0%;  Pred. No. 1.3e-20;
Matches  44;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          ||||||||||||||||||||||||||||||||||||||||||||
Db  303 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 346

RESULT 7
AAB20219
ID   AAB20219 standard; protein; 664 AA.
XX
AC   AAB20219;
XX
DT   14-MAY-2001  (first entry)
XX
DE   Human Chfr (checkpoint with FHA and ring finger) protein.
XX
KW   Checkpoint with forkhead associated domain and ring finger; Chfr; human;
KW   mitosis; cell cycle; tumour; diagnosis; antitumour; drug screening;
KW   ubiquitin-protein ligase.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          31..103
FT                   /label= Forkhead-associated_domain
FT   Domain          303..346
FT                   /label= Ring_finger-domain
FT   Region          476..641
FT                   /note= "cysteine-rich region"
FT   Misc-difference 580
FT                   /note= "Met in U2OS cells"
XX
PN   WO200109150-A2.
XX
PD   08-FEB-2001.
XX
PF   14-JUN-2000; 2000WO-US016391.
XX
PR   29-JUL-1999;   99US-0146194P.
XX
PA   (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI   Halazonetis T, Scolnick D;
XX
```

DE   Human protein sequence SEQ ID NO:12128.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000;   2000JP-00118776.
PR   02-MAY-2000;   2000JP-00183767.
PR   09-JUN-2000;   2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 12128; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises at least 15 nucleotides and the combination of
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 623 AA;

Query Match   100.0%; Score 266; DB 4; Length 623;
Best Local Similarity   100.0%; Pred. No. 1.3e-20;
Matches   44; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
           |||||||||||||||||||||||||||||||||||||||||||
Db   262 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 305

RESULT 4
ABB97233
ID   ABB97233 standard; protein; 623 AA.
XX
AC   ABB97233;
XX
DT   27-JUN-2002 (first entry)

XX
DE   Novel human protein SEQ ID NO: 501.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag.
XX
OS   Homo sapiens.
XX
PN   WO20022660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US026015.
XX
PR   11-SEP-2000; 2000US-00659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   N-PSDB; ABN32419.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis.
XX
PS   Example 2; SEQ ID NO 501; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC   and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC   rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a protein of the invention
XX
SQ   Sequence 623 AA;

Query Match   100.0%; Score 266; DB 5; Length 623;
Best Local Similarity   100.0%; Pred. No. 1.3e-20;
Matches   44; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
           |||||||||||||||||||||||||||||||||||||||||||
Db   262 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 305

RESULT 5
AAB93168
ID   AAB93168 standard; protein; 652 AA.
XX
AC   AAB93168;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human protein sequence SEQ ID NO:12100.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX

```
CC     to prevent skin aging due to sunburn, to maintain organs before
CC     transplantation, for supporting cell culture of primary tissues, to
CC     regenerate tissues and in chemotaxis. The polypeptides can also be used
CC     as a food additive or preservative to increase or decrease storage
CC     capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC     minerals, cofactors and other nutritional components. The present
CC     sequence represents a novel secreted protein of the invention. Note: The
CC     sequence data for this patent did not form part of the printed

Query Match             100.0%;  Score 266; DB 4; Length 426;
Best Local Similarity   100.0%;  Pred. No. 9.2e-21;
Matches    44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1  TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          |||||||||||||||||||||||||||||||||||||||||||||
Db   213  TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 256


RESULT 2
ABU54925
ID   ABU54925 standard; protein; 426 AA.
XX
AC   ABU54925;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polypeptide #12.
XX
KW   Human; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBB/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   N-PSDB; ABX73184.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 11; SEQ ID NO 809; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vesibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC   ABU55748 represent human novel polypeptides of the invention
XX
SQ   Sequence 426 AA;

Query Match             100.0%;  Score 266; DB 6; Length 426;
Best Local Similarity   100.0%;  Pred. No. 9.2e-21;
Matches    44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     1  TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 44
          |||||||||||||||||||||||||||||||||||||||||||||
Db   213  TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 256


RESULT 3
AAB93182
ID   AAB93182 standard; protein; 623 AA.
XX
AC   AAB93182;
XX
DT   26-JUN-2001  (first entry)
XX
```

```
PR   14-AUG-2000;  2000US-0225213P.
PR   14-AUG-2000;  2000US-0225214P.
PR   14-AUG-2000;  2000US-0225266P.
PR   14-AUG-2000;  2000US-0225267P.
PR   14-AUG-2000;  2000US-0225268P.
PR   14-AUG-2000;  2000US-0252270P.
PR   14-AUG-2000;  2000US-0254447P.
PR   14-AUG-2000;  2000US-0257757P.
PR   14-AUG-2000;  2000US-0257758P.
PR   14-AUG-2000;  2000US-0257759P.
PR   18-AUG-2000;  2000US-0262279P.
PR   22-AUG-2000;  2000US-0266681P.
PR   22-AUG-2000;  2000US-0268868P.
PR   23-AUG-2000;  2000US-0289924P.
PR   30-AUG-2000;  2000US-0271182P.
PR   01-SEP-2000;  2000US-0227009P.
PR   01-SEP-2000;  2000US-0292887P.
PR   01-SEP-2000;  2000US-0293343P.
PR   05-SEP-2000;  2000US-0229344P.
PR   05-SEP-2000;  2000US-0239345P.
PR   05-SEP-2000;  2000US-0229509P.
PR   06-SEP-2000;  2000US-0229513P.
PR   06-SEP-2000;  2000US-0230437P.
PR   06-SEP-2000;  2000US-0230438P.
PR   08-SEP-2000;  2000US-0231242P.
PR   08-SEP-2000;  2000US-0231243P.
PR   08-SEP-2000;  2000US-0231244P.
PR   08-SEP-2000;  2000US-0231413P.
PR   08-SEP-2000;  2000US-0231414P.
PR   08-SEP-2000;  2000US-0232080P.
PR   12-SEP-2000;  2000US-0232081P.
PR   14-SEP-2000;  2000US-0231968P.
PR   14-SEP-2000;  2000US-0232397P.
PR   14-SEP-2000;  2000US-0232398P.
PR   14-SEP-2000;  2000US-0232399P.
PR   14-SEP-2000;  2000US-0232400P.
PR   14-SEP-2000;  2000US-0232401P.
PR   14-SEP-2000;  2000US-0233063P.
PR   14-SEP-2000;  2000US-0233064P.
PR   21-SEP-2000;  2000US-0233065P.
PR   21-SEP-2000;  2000US-0234223P.
PR   25-SEP-2000;  2000US-0234274P.
PR   26-SEP-2000;  2000US-0234997P.
PR   27-SEP-2000;  2000US-0234998P.
PR   27-SEP-2000;  2000US-0235484P.
PR   29-SEP-2000;  2000US-0235834P.
PR   29-SEP-2000;  2000US-0235836P.
PR   29-SEP-2000;  2000US-0236327P.
PR   29-SEP-2000;  2000US-0236367P.
PR   29-SEP-2000;  2000US-0236368P.
PR   29-SEP-2000;  2000US-0236369P.
PR   29-SEP-2000;  2000US-0236370P.
PR   02-OCT-2000;  2000US-0236802P.
PR   02-OCT-2000;  2000US-0237037P.
PR   02-OCT-2000;  2000US-0237038P.
PR   02-OCT-2000;  2000US-0237039P.
PR   02-OCT-2000;  2000US-0237040P.
PR   13-OCT-2000;  2000US-0239935P.
PR   13-OCT-2000;  2000US-0239937P.
PR   20-OCT-2000;  2000US-0240960P.
PR   20-OCT-2000;  2000US-0241221P.
PR   20-OCT-2000;  2000US-0241785P.
PR   20-OCT-2000;  2000US-0241786P.
PR   20-OCT-2000;  2000US-0241787P.
PR   20-OCT-2000;  2000US-0241808P.
PR   20-OCT-2000;  2000US-0241809P.
PR   20-OCT-2000;  2000US-0241826P.
PR   01-NOV-2000;  2000US-0246474P.
PR   08-NOV-2000;  2000US-0246475P.
PR   08-NOV-2000;  2000US-0246476P.
PR   08-NOV-2000;  2000US-0246477P.
PR   08-NOV-2000;  2000US-0246478P.

PR   08-NOV-2000;  2000US-0245523P.
PR   08-NOV-2000;  2000US-0246524P.
PR   08-NOV-2000;  2000US-0245525P.
PR   08-NOV-2000;  2000US-0245526P.
PR   08-NOV-2000;  2000US-0246527P.
PR   08-NOV-2000;  2000US-0246528P.
PR   08-NOV-2000;  2000US-0246532P.
PR   08-NOV-2000;  2000US-0246609P.
PR   08-NOV-2000;  2000US-0246610P.
PR   08-NOV-2000;  2000US-0246611P.
PR   08-NOV-2000;  2000US-0246613P.
PR   17-NOV-2000;  2000US-0249207P.
PR   17-NOV-2000;  2000US-0249208P.
PR   17-NOV-2000;  2000US-0249209P.
PR   17-NOV-2000;  2000US-0249210P.
PR   17-NOV-2000;  2000US-0249211P.
PR   17-NOV-2000;  2000US-0249212P.
PR   17-NOV-2000;  2000US-0249213P.
PR   17-NOV-2000;  2000US-0249214P.
PR   17-NOV-2000;  2000US-0249215P.
PR   17-NOV-2000;  2000US-0249216P.
PR   17-NOV-2000;  2000US-0249217P.
PR   17-NOV-2000;  2000US-0249218P.
PR   17-NOV-2000;  2000US-0249244P.
PR   17-NOV-2000;  2000US-0249245P.
PR   17-NOV-2000;  2000US-0249264P.
PR   17-NOV-2000;  2000US-0249265P.
PR   17-NOV-2000;  2000US-0249297P.
PR   17-NOV-2000;  2000US-0249299P.
PR   17-NOV-2000;  2000US-0249300P.
PR   01-DEC-2000;  2000US-0250160P.
PR   01-DEC-2000;  2000US-0250391P.
PR   05-DEC-2000;  2000US-0251030P.
PR   05-DEC-2000;  2000US-0251988P.
PR   05-DEC-2000;  2000US-0256719P.
PR   06-DEC-2000;  2000US-0251479P.
PR   08-DEC-2000;  2000US-0251856P.
PR   08-DEC-2000;  2000US-0251868P.
PR   08-DEC-2000;  2000US-0251869P.
PR   08-DEC-2000;  2000US-0251989P.
PR   08-DEC-2000;  2000US-0251990P.
PR   11-DEC-2000;  2000US-0254097P.
PR   05-JAN-2001;  2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   N-PSDB; AAS25843.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 11; SEQ ID NO 809; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      May  7, 2004, 14:36:57 ; Search time 9.01373 Seconds
                              (without alignments)
                          1379.240 Million cell updates/sec

Title:        US-10-048-046-2_COPY_303_346
Perfect score: 266
Sequence:     1 TCIICQDLLHDCVSLQPCMH.........ACYSGWMERSSLCPTCRCPV 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:     1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:    1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    A_Geneseq_29Jan04:*
              1: geneseqp1980s:*
              2: geneseqp1990s:*
              3: geneseqp2000s:*
              4: geneseqp2001s:*
              5: geneseqp2002s:*
              6: geneseqp2003as:*
              7: geneseqp2003bs:*
              8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

```
                       %
Result         Query
No.    Score   Match  Length DB  ID          Description
------ -----   -----  ------ --  ----------  -----------
  1     266    100.0   426   4   AAU15856    Human nov
  2     266    100.0   426   6   ABU54925    Human nov
  3     266    100.0   623   4   AAB93182    Human pro
  4     266    100.0   623   5   ABB97233    Novel hum
  5     266    100.0   652   4   AAB93168    Human pro
  6     266    100.0   664   4   AAB83843    Amino aci
  7     266    100.0   664   4   AAB20219    Human Chf
  8     176     66.2   230   6   AAU16317    Human nov
  9     176     66.2   230   6   ABU55386    Human nov
 10     135     50.8   350   3   AAG20354    Arabidops
 11     135     50.8   350   3   AAG37660    Arabidops
 12     135     50.8   426   3   AAG20353    Arabidops
 13     135     50.8   426   3   AAG37659    Arabidops
 14     135     50.8   453   3   AAG20352    Arabidops
 15     135     50.8   453   3   AAG37658    Arabidops
 16     129     48.5    67   2   AAR60624    Varicella
 17     115     43.2   242   4   AAM39926    Human pol
 18     115     43.2   247   8   ADE77120    Human pro
 19     115     43.2   254   5   ABB79353    Human ova
 20     115     43.2   261   4   AAM41712    Human pol
 21     113.5   42.7   251   3   AAG06408    Arabidops
 22     113.5   42.7   322   3   AAG06407    Arabidops
 23     113.5   42.7   426   3   AAG06406    Arabidops
 24     112     42.1   242   4   AAM79067    Human pro
 25     111     41.7   495   6   ABR56805    Human sec
```

```
 26     110     41.4   498   2   AAW80439    Feline he
 27     108.5   40.8   485   4   AAB95017    Human pro
 28     107     40.2   551   2   AAW27150    HMG-CoA r
 29     104     39.1   330   4   AAE01417    Human sec
 30     103     38.7   328   4   AAE01418    Human sec
 31     103     38.7   489   5   ABG64061    Human alb
 32     103     38.7   505   4   AAE01380    Human gen
 33     103     38.7   616   5   ABB83078    Synovioli
 34     103     38.7   617   5   AAE01355    Human gen
 35     103     38.7   617   5   ABG79641    Human nov
 36     103     38.7   617   5   ABG64063    Human alb
 37     103     38.7   617   6   ABB83074    Synovioli
 38     103     38.7   617   6   AAE37275    Human gen
 39     102     38.3   328   4   ABB64164    Drosophil
 40     101     38.0   568   4   AAB18230    Plasmodiu
 41     100.5   37.8   499   5   AAO22598    499-mer S
 42     100.5   37.8  1208   4   AAB95292    Human pro
 43     100.5   37.8  1213   5   AAO22597    1213-mer
 44     100     37.6   222   4   ABB60472    Drosophil
 45     100     37.6   356   6   ABU65227    Novel hum
```

```
                              ALIGNMENTS

RESULT 1
AAU15856
ID   AAU15856 standard; protein; 426 AA.
XX
AC   AAU15856;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 809.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; antiarthritic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmological; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
```

RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Colon;
RA   Strausberg R.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF076687; AAD20943.1; -.
DR   EMBL; BC055061; AAH55061.1; -.
DR   MGD; MGI:1351625; Nbn.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   InterPro; IPR001357; BRCT.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00533; BRCT; 1.
DR   Pfam; PF00498; FHA; 1.
DR   SMART; SM00292; BRCT; 1.
DR   SMART; SM00240; FHA; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   751 AA;   83794 MW;   C9F597CC08227B2C CRC64;

Query Match   23.0%; Score 87; DB 11; Length 751;
Best Local Similarity   26.6%; Pred. No. 0.15;
Matches 21; Conservative 21; Mismatches 27; Indels 10; Gaps   2;

Qy   2 LLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVD-------EKSGQVTLEDTSTSGT 52
          || ||:|::| ||::|: :|||:::|:| :|:   :: :|::|| |||||
Db   18 LLAGVEVVVG-RKNCGILIENDQSISRNHAVLTVNFPVTSLSQTDEIPTLTIKDNSKYGT 76

Qy   53 VINKLKVVKKQTCPLQTGD 71
          :|: | :|||||
Db   77 FVNEEKMQTGLSCTLKTGD 95

RESULT 15
Q8G6Q4
ID   Q8G6Q4   PRELIMINARY;   PRT;   176 AA.
AC   Q8G6Q4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein with FHA domain.
GN   BL0584.
OS   Bifidobacterium longum.
OC   Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC   Bifidobacteriaceae; Bifidobacterium.
OX   NCBI_TaxID=216816;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCC 2705;
RX   MEDLINE=22294977; PubMed=12381787;
RA   Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA   Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,

RA   Pridmore R.D., Arigoni F.;
RT   "The genome sequence of Bifidobacterium longum reflects its adaptation
RT   to the human gastrointestinal tract.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR   EMBL; AE014680; AAN24408.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00498; FHA; 1.
DR   SMART; SM00240; FHA; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   176 AA;   19530 MW;   1612CDF8C9A93CF6 CRC64;

Query Match   22.8%; Score 86; DB 16; Length 176;
Best Local Similarity   32.8%; Pred. No. 0.035;
Matches 22; Conservative 15; Mismatches 28; Indels   2; Gaps   2;

Qy   1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLED-TSTSGTVINKLKV 59
          ||| ||:|::| |:: |:|| | | :| :|:||:|:||| |: |: || :: ::
Db   96 VPLABADITLGRAASNTVVL-DDEFVSSHHARVYRDTRSGQWAIEDLNSTNGTVVNQQRI 154

Qy   60 VKKQTCP 66
          | |
Db   155 NRPTILP 161

Search completed: May 7, 2004, 14:49:02
Job time : 11.5607 secs

```
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain stem;
RA   Mas C., Bourgeois F., Simonneau M.;
RT   "Isolation of 50 cDNAs differentially expressed in embryonic forebrain
RT   as compared to mid and hindbrain: a strategy to identify candidate
RT   genes involved in human neurodevelopmental diseases.";
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF092840; AAC62113.1; -.
DR   MGD; MGI:1351625; Nbn.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   InterPro; IPR001357; BRCT.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00533; BRCT; 1.
DR   SMART; SM00292; BRCT; 1.
DR   SMART; SM00240; FHA; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   SEQUENCE   751 AA;  83826 MW;  9D136BBC23DC51D9 CRC64;

Query Match             23.0%;  Score 87;  DB 11;  Length 751;
Best Local Similarity   26.6%;  Pred. No. 0.15;  -
Matches      21; Conservative   21; Mismatches   27; Indels   10; Gaps    2;

Qy    2 LLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVD--------EKSGQVTLEDTSTSGT 52
        ||  |: |: |:       :: :: :|: |:     :: ::||
Db   18 LLAGVEYVVG-RKNCGILIENDQSISRNHAVLTVNFPVTSLSQTDEIPTLTIKDNSKYGT 76

Qy   53 VINKLKVVKKQTCPLQTGD 71
        :|: |:     :||:|||
Db   77 FVNEEKMQTGLSCTLKTGD 95

RESULT 12
Q9R1X1
ID   Q9R1X1;
AC   Q9R1X1;                PRELIMINARY;       PRT;     751 AA.
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nibrin.
GN   NBN OR NBS1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Testis;
RA   Saito T.;
RT   "Structure of the mouse Nijmegen breakage syndrome (Nibrin/Nbs1)
RT   protein.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016988; BAA76298.1; -.
DR   MGD; MGI:1351625; Nbn.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   InterPro; IPR001357; BRCT.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00533; BRCT; 1.
DR   Pfam; PF00498; FHA; 1.
DR   SMART; SM00292; BRCT; 1.
DR   SMART; SM00240; FHA; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   SEQUENCE   751 AA;  83842 MW;  AEA75A61F57838F9 CRC64;

Query Match             23.0%;  Score 87;  DB 11;  Length 751;
Best Local Similarity   26.6%;  Pred. No. 0.15;
```

```
Matches      21; Conservative   21; Mismatches   27; Indels   10; Gaps    2;

Qy    2 LLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVD--------EKSGQVTLEDTSTSGT 52
        ||  |: |: |:       :: :: :|: |:     :: ::||
Db   18 LLAGVEYVVG-RKNCGILIENDQSISRNHAVLTVNFPVTSLSQTDEIPTLTIKDNSKYGT 76

Qy   53 VINKLKVVKKQTCPLQTGD 71
        :|: |:     :||:|||
Db   77 FVNEEKMQTGLSCTLKTGD 95

RESULT 13
Q81II6
ID   Q81II6;
AC   Q81II6;                PRELIMINARY;       PRT;     751 AA.
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nibrin.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CZECH II;
RA   Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC044773; AAH44773.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00498; FHA; 1.
DR   SMART; SM00240; FHA; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   SEQUENCE   751 AA;  83829 MW;  DD78D11F3C1AE3F5 CRC64;

Query Match             23.0%;  Score 87;  DB 11;  Length 751;
Best Local Similarity   26.6%;  Pred. No. 0.15;  -
Matches      21; Conservative   21; Mismatches   27; Indels   10; Gaps    2;

Qy    2 LLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVD--------EKSGQVTLEDTSTSGT 52
        ||  |: |: |:       :: :: :|: |:     :: ::||
Db   18 LLAGVEYVVG-RKNCGILIENDQSISRNHAVLTVNFPVTSLSQTDEIPTLTIKDNSKYGT 76

Qy   53 VINKLKVVKKQTCPLQTGD 71
        :|: |:     :||:|||
Db   77 FVNEEKMQTGLSCTLKTGD 95

RESULT 14
Q9R207
ID   Q9R207;
AC   Q9R207;                PRELIMINARY;       PRT;     751 AA.
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nibrin (Hypothetical protein).
GN   NBN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20108791; PubMed=10640816;
RA   Vissinga C.S., Yeo T.C., Woessner J., Massa H.F., Wilson R.K.,
RA   Trask B.J., Concannon P.;
RT   "Identification, characterization, and mapping of a mouse homolog of
RT   the gene mutated in Nijmegen breakage syndrome.";
RL   Cytogenet. Cell Genet. 87:80-84(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Colon;
```

```
ID   Q87PU0          PRELIMINARY;      PRT;   504 AA.
AC   Q87PU0;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   VP1411.
OS   Vibrio parahaemolyticus.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=670;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIMD 2210633 / Serotype O3:K6;
RX   MEDLINE=22508454; PubMed=12620739;
RA   Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA   Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Iida T.,
RA   Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT   "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT   distinct from that of V. cholerae.";
RL   Lancet 361:743-749(2003).
DR   EMBL; AP005078; BAC59674.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00498; FHA; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   504 AA;  55469 MW;  2F96C7SA4F83648E CRC64;

Query Match          28.4%;
Best Local Similarity 36.4%;
Matches  24;  Conservative  12;  Mismatches  27;  Indels   3;  Gaps   2;

Qy          9 TIGRRRGCDLSFPS-NKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPL 67
              ::||  |:.:|||| :||||||:|| |:::|:|.||:|.|::|| |||:||.||
Db         31 SVGRAPSCTLSLTDHNRFISGTHCLISV--YGDTPYISDVSTNGTMVNGNKILKNQPISI 88

Qy         68 QTGDVI 73
              |||:
Db         89 VEGDVV 94

RESULT 9
ID   Q97LQ0          PRELIMINARY;      PRT;   159 AA.
AC   Q97LQ0;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   FHA-domain containing secreted protein.
GN   CAC0504.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21159325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007565; AAK78484.1; -.
DR   PIR; A96962; A96962.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00498; FHA; 1.
DR   SMART; SM00240; FHA; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
KW   Complete proteome.
SQ   SEQUENCE   159 AA;  17667 MW;  F4F44A3D3BF05AA1 CRC64;

Query Match          23.3%;           Score 88;  DB 16;  Length 159;
Best Local Similarity 41.4%;           Pred. No. 0.018;
Matches  29;  Conservative  12;  Mismatches  21;  Indels   8;  Gaps   5;

Qy          6 REWTIGRRRGCDLSFPSNK-LVSGDHCRIVDEKSGQVTLED-TSTSGTVINKLKVVKKQ 63
              |:||||||||||||||:| :::.:||| |||:.||||.||  ||.||:|||::|||:
Db         85 REITIGRKD--DNSIMLNEGYVSGHHARVYL--RNNQYILEDLNSTNGTVLNGQKI--KS 138

Qy         64 TCPLQTGDVI 73
              ::||:||
Db        139 KAYIKSGDEI 148

RESULT 10
ID   Q8CCY0          PRELIMINARY;      PRT;   546 AA.
AC   Q8CCY0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nibrin (Fragment).
GN   NBN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Medulla oblongata;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK031933; BAC27610.1; -.
DR   MGD; MGI:1351625; Nbn.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   InterPro; IPR001357; BRCT.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00533; BRCT; 1.
DR   Pfam; PF00498; FHA; 1.
DR   SMART; SM00240; FHA; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
FT   NON_TER       546    546
SQ   SEQUENCE   546 AA;  60269 MW;  69394A2537EFBF20 CRC64;

Query Match          23.0%;           Score 87;  DB 11;  Length 546;
Best Local Similarity 26.6%;           Pred. No. 0.1;
Matches  21;  Conservative  21;  Mismatches  27;  Indels  10;  Gaps   2;

Qy          2 LLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVD--------EKSGQVTLEDTSTSGT 52
              || |: :|| : :|| :: :||||||.|||||.|        :::.||:||||||||
Db         18 LLAGVEYVVG-RKNCGILENDQSISRNHAVLIVNFPVTSLSQTDEIPLTIKDNSKYGT 76

Qy         53 VINKLKVVKKQTCPLQTGD 71
              :||:||||||||||||||
Db         77 FVNEEKMQTGLSCTLKTGD 95

RESULT 11
ID   O88981          PRELIMINARY;      PRT;   751 AA.
AC   O88981;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nibrin.
GN   NBN.
```

```
Db        31 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 90

Qy        61 KKQTCPLQTGDVI 73
             |||||||||||||
Db        91 KKQTCPLQTGDVI 103

RESULT 5
Q8BWH4
ID   Q8BWH4;            PRELIMINARY;       PRT;    254 AA.
AC   Q8BWH4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cell cycle checkpoint protein CHFR (Fragment).
GN   5730484M20RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Lung;
RX   MEDLINE=22354683; PubMed=12466851;
RA   the FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK052473; BAC35008.1; -.
DR   MGD; MGI:2444898; 5730484M20Rik.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00498; FHA; 1.
DR   SMART; SM00240; FHA; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
FT   NON_TER      254    254
SQ   SEQUENCE   254 AA;  27819 MW;  0340571A2BAEB6EF CRC64;

Query Match             86.2%;        Score 326;  DB 11;  Length 254;
Best Local Similarity   83.6%;        Pred. No. 1.6e-30;
Matches    61;  Conservative    8;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
             :|| |||||||||||||||||||||||||:|| ::|||::|||||||||:|||||:||
Db        31 ILLWKREWTIGRRRGCDLSFPSNKLVSGDHCKLTVDEISGEVTLEDTSTNGTVINKLQVV 90

Qy        61 KKQTCPLQTGDVI 73
             |||| ||:|:||
Db        91 KKQTYPLQSGDII 103

RESULT 6
Q8I0L3
ID   Q8I0L3;            PRELIMINARY;       PRT;    664 AA.
AC   Q8I0L3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RIKEN cDNA 5730484M20 gene.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090; .
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Limb;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC049792; AAH49792.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_Ring.
```

```
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 2.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE   664 AA;  73871 MW;  D651BE3B463DBB6 CRC64;

Query Match             86.2%;        Score 326;  DB 11;  Length 664;
Best Local Similarity   83.6%;        Pred. No. 4.9e-30;
Matches    61;  Conservative    8;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
             :|| |||||||||||||||||||||||||:|| ::|||::|||||||||:|||||:||
Db        31 ILLWKREWTIGRRRGCDLSFPSNKLVSGDHCKLTVDEISGEVTLEDTSTNGTVINKLQVV 90

Qy        61 KKQTCPLQTGDVI 73
             |||| ||:|:||
Db        91 KKQTYPLQSGDII 103

RESULT 7
Q8BJZ9
ID   Q8BJZ9;            PRELIMINARY;       PRT;    663 AA.
AC   Q8BJZ9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cell cycle checkpoint protein CHFR.
GN   5730484M20RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Body;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK077629; BAC36912.1; -.
DR   MGD; MGI:2444898; 5730484M20Rik.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_Ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 2.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE   663 AA;  73855 MW;  42BD653DCB4ADF45 CRC64;

Query Match             84.9%;        Score 321;  DB 11;  Length 663;
Best Local Similarity   82.2%;        Pred. No. 1.9e-29;
Matches    60;  Conservative    8;  Mismatches    5;  Indels    0;  Gaps    0;

Qy         1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
             :|| |||||||||||||||||||||||||:|| ::|||::|||||||||:|||||:||
Db        31 ILLWKREWTIGRRRGCDLSFPSNKLVSGDHCKLTVDEISGEVTLEDTSTNGTVINKLQVV 90

Qy        61 KKQTCPLQTGDVI 73
             |||| ||:|||
Db        91 KKQTYPLHSGDII 103

RESULT 8
Q87PU0
```

Qy   1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 90

Qy   61 KKQTCPLQTGDVI 73
        |||||||||||||
Db   91 KKQTCPLQTGDVI 103

RESULT 2
Q96SL3
ID   Q96SL3   PRELIMINARY;   PRT;   652 AA.
AC   Q96SL3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ14781.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AK027687; BAB55297.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_Ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 652 AA; 72030 MW; 572F2CE6D1743D80 CRC64;

Query Match   100.0%; Score 378; DB 4; Length 652;
Best Local Similarity   100.0%; Pred. No. 3.1e-36;
Matches   73; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 90

Qy   61 KKQTCPLQTGDVI 73
        |||||||||||||
Db   91 KKQTCPLQTGDVI 103

RESULT 3
Q96EP1
ID   Q96EP1   PRELIMINARY;   PRT;   652 AA.
AC   Q96EP1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Strausberg R.;

RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; BC012072; AAH12072.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_Ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 652 AA; 72058 MW; 1A50773CEA679F07 CRC64;

Query Match   100.0%; Score 378; DB 4; Length 652;
Best Local Similarity   100.0%; Pred. No. 3.1e-36;
Matches   73; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 90

Qy   61 KKQTCPLQTGDVI 73
        |||||||||||||
Db   91 KKQTCPLQTGDVI 103

RESULT 4
Q9NRT4
ID   Q9NRT4   PRELIMINARY;   PRT;   664 AA.
AC   Q9NRT4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cell cycle checkpoint protein CHFR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20388685; PubMed=10935642;
RA   Scolnick D.M., Halazonetis T.D.;
RT   "Chfr defines a mitotic stress checkpoint that delays entry into
RT   metaphase.";
RL   Nature 406:430-435(2000).
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AF170724; AAF91084.1; -.
DR   PDB; 1LGP; 07-AUG-02.
DR   PDB; 1LGQ; 07-AUG-02.
DR   Genew; HGNC:20455; CHFR.
DR   GO; GO:0007093; P:mitotic checkpoint; TAS.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_Ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 664 AA; 73386 MW; 141A1E7EEFAE36A2 CRC64;

Query Match   100.0%; Score 378; DB 4; Length 664;
Best Local Similarity   100.0%; Pred. No. 3.2e-36;
Matches   73; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
              GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     May 7, 2004, 14:40:47 ; Search time 10.5607 Seconds
                                     (without alignments)
                                     2180.991 Million cell updates/sec

Title:        US-10-048-046-2_COPY_31_103
Perfect score: 378
Sequence:     1 VLLRKREWTIGRRRGCDLSF.........INKLIKVVKKQTCPLQTGDVI 73

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:     1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:     1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    .SPTREMBL_25:*
          1:  sp_archea:*
          2:  sp_bacteria:*
          3:  sp_fungi:*
          4:  sp_human:*
          5:  sp_invertebrate:*
          6:  sp_mammal:*
          7:  sp_mhc:*
          8:  sp_organelle:*
          9:  sp_phage:*
         10:  sp_plant:*
         11:  sp_rodent:*
         12:  sp_virus:*
         13:  sp_vertebrate:*
         14:  sp_unclassified:*
         15:  sp_rvirus:*
         16:  sp_bacteriap:*
         17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	378	100.0	623	4	Q9NVD5	Q9nvd5 homo sapien
2	378	100.0	652	4	Q96SL3	Q96sl3 homo sapien
3	378	100.0	652	4	Q96EP1	Q96ep1 homo sapien
4	378	100.0	664	4	Q9NRT4	Q9nrt4 homo sapien
5	326	86.2	254	11	Q8BWH4	Q8bwh4 mus musculu
6	326	86.2	664	11	Q8I0L3	Q8i0l3 mus musculu
7	321	84.9	663	11	Q8BJ29	Q8bj29 mus musculu
8	107.5	28.4	504	16	Q87PU0	Q87pu0 vibrio para
9	88	23.3	159	16	Q97LQ0	Q97lq0 clostridium
10	87	23.0	546	11	Q8CCY0	Q8ccy0 mus musculu
11	87	23.0	751	11	Q88981	Q88981 mus musculu
12	87	23.0	751	11	Q9R1X1	Q9r1x1 mus musculu
13	87	23.0	751	11	Q81I6	Q81i6 mus musculu
14	87	23.0	751	11	Q9R207	Q9r207 mus musculu
15	86	22.8	176	16	Q8G6Q4	Q8g6q4 bifidobacte
16	83	22.0	243	16	Q7UWU2	Q7uwu2 rhodopirell
17	82.5	21.8	729	10	Q9FMB8	Q9fmb8 arabidopsis
18	79.5	21.0	497	16	Q9I751	Q9i751 pseudomonas
19	79.5	21.0	1236	5	Q9NED7	Q9ned7 leishmania
20	79.5	21.0	250	16	Q7UGY0	Q7ugy0 rhodopirell
21	79	20.9	345	4	Q9UBH7	Q9ubh7 homo sapien
22	79	20.9	345	4	Q9Y242	Q9y242 homo sapien
23	79	20.9	345	4	Q9BUM2	Q9bum2 homo sapien
24	79	20.9	345	6	Q7YR48	Q7yr48 pan troglod
25	79	20.9	359	4	Q13176	Q13176 homo sapien
26	79	20.9	359	16	Q9I1K6	Q9i1k6 rhizobium l
27	77.5	20.5	460	16	Q89P85	Q89p85 bradyrhizob
28	77.5	20.5	528	3	Q874N8	Q874n8 candida gla
29	76.5	20.2	517	13	Q9I8V3	Q9i8v3 xenopus lae
30	76	20.1	2541	5	Q19663	Q19663 caenorhabdi
31	74.5	19.7	263	11	Q99KJ5	Q99kj5 mus musculu
32	74.5	19.7	263	11	Q8BPA3	Q8bpa3 mus musculu
33	74.5	19.7	517	13	Q98TW0	Q98tw0 xenopus lae
34	74.5	19.7	689	10	Q9FNZ7	Q9fnz7 oryza sativ
35	73.5	19.4	526	3	Q874M7	Q874m7 kluyveromyc
36	73	19.3	144	16	Q890Y0	Q890y0 clostridium
37	73	19.3	503	13	Q902Y5	Q902y5 brachydanio
38	71.5	18.9	398	16	Q87HC2	Q87hc2 vibrio para
39	71	18.8	503	13	Q803B2	Q803e2 brachydanio
40	70.5	18.7	168	5	Q86DY9	Q86dy9 schistosoma
41	70.5	18.7	230	16	Q9KBZ7	Q9kbz7 bacillus ha
42	70.5	18.7	301	2	O74388	O74388 schizosacch
43	70	18.5	254	2	Q9S475	Q9s475 myxococcus
44	70	18.5	388	2	O52176	O52176 myxococcus
45	68.5	18.1	157	16	P74513	P74513 synechocyst

ALIGNMENTS

```
RESULT 1
Q9NVD5
ID   Q9NVD5        PRELIMINARY;       PRT;     623 AA.
AC   Q9NVD5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ10796.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Nagai S., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AK001658; BAA91817.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 623 AA;   69204 MW;   45186D33DAE52711 CRC64;

Query Match              100.0%;   Score 378;   DB 4;   Length 623;
Best Local Similarity    100.0%;   Pred. No. 3e-36;
Matches   73;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;
```

```
Qy      49  T-----SGTVI-----NKLKVVKKQTC 65
            |   ||.|  . :           ||||
Db     129  TNHFEIAGTTVKSGWFKSTTYITLDQTC 156

RESULT 14
TRMU_CORGL
ID   TRMU_CORGL     STANDARD;      PRT;    365 AA.
AC   Q8NR24;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Probable tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase
DE   (EC 2.1.1.61).
GN   TRMU OR CGL1240.
OS   Corynebacterium glutamicum (Brevibacterium flavum).
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=1718;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA   Nakagawa S.;
RT   "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC       homocysteine + tRNA containing 5-methylaminomethyl-2-
CC       thiouridylate.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Belongs to the trmU family.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on  its
CC   use  by non-profit institutions as long  as its content  is in  no  way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AP005277; BAB98633.1; -.
DR   HAMAP; MF_00144; -; 1.
DR   InterPro; IPR004506; TrmU.
DR   Pfam; PF03054; tRNA_Me_trans; 1.
DR   TIGRFAMs; TIGR00420; trmU; 1.
KW   Transferase; Methyltransferase; tRNA processing; Complete proteome.
SQ   SEQUENCE 365 AA;  39237 MW;  21F8EE1B2B3BDC63 CRC64;

Query Match           16.0%;    Score 60.5;  DB 1;   Length 365;
Best Local Similarity 29.6%;    Pred. No.18;
Matches 21;   Conservative 14;   Mismatches 23;  Indels 13;  Gaps 4;

Qy    7  EWTIGRRGCDLSFPSNKLVSGDHCRIV--VDEKSGQVTL--EDTSTSGTVINKLKVVK 61
            |:|||:||:|:  |::  ||  |||  :||  ::|||
Db  237  EFTIGQRKGLDIKAPA----ADGRPRYVTDIDAKTGTVTVGTRENLKISTIHADRLKFLH 292

Qy   62  KQTCPLQTGDV 72
            |:
Db  293  ----PAMDGQI 299

RESULT 15
Y634_CHLMU
ID   Y634_CHLMU     STANDARD;      PRT;    353 AA.
AC   Q9PK39;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein TC0634 precursor.
GN   TC0634.
OS   Chlamydia muridarum.
OC   Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=83560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MoPn / Nigg;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
CC   -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN1058/CT355/TC0634
CC       FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on  its
CC   use  by non-profit institutions as long  as its content  is in  no  way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AE002332; AAF39463.1; -.
DR   PIR; E81680; E81680.
DR   TIGR; TC0634; -.
KW   Hypothetical protein; Signal; Complete proteome.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    353       HYPOTHETICAL PROTEIN TC0634.
SQ   SEQUENCE 353 AA;  39840 MW;  3048C59F5B7BB90E CRC64;

Query Match           15.9%;    Score 60;  DB 1;   Length 353;
Best Local Similarity 32.2%;    Pred. No. 20;
Matches 19;   Conservative 10;   Mismatches 16;  Indels 14;  Gaps 3;

Qy   15  GCDLSFPSNKLVSGD----HCRIV-VDEKSGQVTLEDTSTSGTVINKLKVVKKQTCP 66
            :|:||||:|          ||||  :|:|  :||
Db   24  GYEGSFPSSSLEQNPSGVAIHNRVLFKIDEDTVVTLD------VIHKLNILFYSTCP 75

Search completed: May 7, 2004, 14:46:32
Job time : 4.46674 secs
```

```
Matches  22;  Conservative   9;  Mismatches  34;  Indels  12;  Gaps   3;

Qy    8 WTIGRRRGCDLSFPSNKL--------VSGDHCRIVVD----EKSGQVTLEDTSTSGTVINK 56
          ||  : ||             :| |||||||       : |:  |||
Db  118 W-FGRDKSCEYCFDGPLLRRTDKYRTYSKGHFRIFREMGPXNCYIVYIEDHSGNGTFVNT 176

Qy   57 LKVVKKQTCPLQTGDVI 73
          :  :: ||| :  : |
Db  177 ELIGKGKRCPLSNNSEI 193

RESULT 12
VGLM_BUNYW
ID   VGLM_BUNYW     STANDARD;    PRT;   1433 AA.
AC   P04505;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   M polyprotein precursor [Contains: Glycoprotein G2; Nonstructural
DE   protein NS-M; Glycoprotein G1].
GN   M.
OS   Bunyamwera virus.
OC   Viruses; ssRNA negative-strand viruses; Bunyaviridae; Orthobunyavirus.
OX   NCBI_TaxID=35304;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86098655; PubMed=3753629;
RA   Lees J.F., Pringle C.R., Elliott R.M.;
RT   "Nucleotide sequence of the Bunyamwera virus M RNA segment:
RT   conservation of structural features in the Bunyavirus glycoprotein
RT   gene product.";
RL   Virology 148:1-14(1986).
CC   -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins
CC       including nonstructural protein NS-M, glycoprotein G1, and
CC       glycoprotein G2.
CC   -!- SIMILARITY: Belongs to the bunyaviruses M polyprotein family.
```

```
DR   EMBL; M11852; AAA42777.1; -.
DR   PIR; A04101; GNVUBW.
DR   InterPro; IPR005167; Bunya_G1.
DR   InterPro; IPR005168; Bunya_G2.
DR   Pfam; PF03557; Bunya_G1; 1.
DR   Pfam; PF03563; Bunya_G2; 1.
KW   Polyprotein; Glycoprotein; Transmembrane; Nonstructural protein;
KW   Signal.
FT   SIGNAL        1     16
FT   CHAIN        17   1433       M POLYPROTEIN.
FT   CHAIN        17    302       GLYCOPROTEIN G2.
FT   CHAIN       303    477       NONSTRUCTURAL PROTEIN NS-M.
FT   CHAIN       478   1433       GLYCOPROTEIN G1.
FT   TRANSMEM   1382   1409       POTENTIAL.
FT   CARBOHYD     60     60       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    248    248       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    624    624       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1169   1169       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   1433 AA;  162077 MW;  CD61ABDE782018E0 CRC64;

Query Match             16.1%;   Score 61;  DB 1;  Length 1433;
Best Local Similarity   23.9%;   Pred. No. 70;
Matches   21;  Conservative   9;  Mismatches   9;  Indels  30;  Gaps   3;

Qy    6 REWTIGRRGC-----------DLSFPSNKLVSGDHCRIVVDEKSGQV------T 43
          | :|:                :|: :||: ||||| :  :
Db   72 RHWTITDWKACNPVVTAGGSINVIEVDKNLNLVTRNYVCTGDCTIVDRKNAQIIFQTDK 131
```

```
Qy   44 LEDTSTSGTVI-----NKLKVVKKQTC 65
          |||| : |||    : |: |  |||
Db  132 LNHFEVTGTTISTGWFKSKASVTLDRTC 159

RESULT 13
VGLM_BUNL7
ID   VGLM_BUNL7     STANDARD;    PRT;   1441 AA.
AC   P09612;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   M polyprotein precursor [Contains: Glycoprotein G2; Nonstructural
DE   protein NS-M; Glycoprotein G1].
GN   M.
OS   Bunyavirus La Crosse (isolate L74).
OC   Viruses; ssRNA negative-strand viruses; Bunyaviridae; Orthobunyavirus.
OX   NCBI_TaxID=11578;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88089508; PubMed=3694177;
RA   Grady L.J., Sanders M.L., Campbell W.P.;
RT   "The sequence of the M RNA of an isolate of La Crosse virus.";
RL   J. Gen. Virol. 68:3057-3071(1987).
RN   [2]
RP   SEQUENCE OF 1-46 FROM N.A.
RX   MEDLINE=82216937; PubMed=7086954;
RA   Clerx-Van Haaster C.M., Akashi H., Auperin D.D., Bishop D.H.L.;
RT   "Nucleotide sequence analyses and predicted coding of bunyavirus
RT   genome RNA species.";
RL   J. Virol. 41:119-128(1982).
CC   -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins
CC       including nonstructural protein NS-M, glycoprotein G1, and
CC       glycoprotein G2.
CC   -!- SIMILARITY: Belongs to the bunyaviruses M polyprotein family.
```

```
DR   EMBL; D10370; BAA01201.1; -.
DR   EMBL; J02231; -; NOT_ANNOTATED_CDS.
DR   PIR; A29377; GNVULC.
DR   InterPro; IPR005167; Bunya_G1.
DR   InterPro; IPR005168; Bunya_G2.
DR   Pfam; PF03557; Bunya_G1; 1.
DR   Pfam; PF03563; Bunya_G2; 1.
KW   Polyprotein; Glycoprotein; Transmembrane; Nonstructural protein;
KW   Signal.
FT   SIGNAL        1     13
FT   CHAIN        14   1441       M POLYPROTEIN.
FT   CHAIN        22    299       GLYCOPROTEIN G2.
FT   CHAIN       300    473       NONSTRUCTURAL PROTEIN NS-M.
FT   CHAIN       474   1441       GLYCOPROTEIN G1.
FT   CARBOHYD     57     57       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    245    245       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    490    490       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   1177   1177       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT     40     40       C -> S (IN REF. 2).
SQ   SEQUENCE   1441 AA;  162540 MW;  50973CE30973C55B CRC64;

Query Match             16.1%;   Score 61;  DB 1;  Length 1441;
Best Local Similarity   22.7%;   Pred. No. 70;
Matches   20;  Conservative  12;  Mismatches  28;  Indels  28;  Gaps   3;

Qy    6 REWTIGRRGC-----------DLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTS 48
          | |:                ||||||
Db   69 RKWLVSDWHDCRPKKIVGGHINVIEVGDDLSLHTESYVCSADCTIGVDKETAQVRLQTDT 128
```

CC   -!- SIMILARITY: Contains 1 fork-head domain.
CC   -!- SIMILARITY: Contains 1 FHA domain.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; L38848; AAA60938.1; -.
DR   EMBL; Z38059; CAA86147.1; -.
DR   PIR; S48403; S48403.
DR   HSSP; Q63245; 2HFH.
DR   GermOnline; 139666; -.
DR   TRANSFAC; T03303; -.
DR   SGD; S0001193; FKH1.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0006347; P:chromatin silencing at HML and HMR (sensu S. ..; IGI.
DR   GO; GO:0007124; P:pseudohyphal growth; IGI.
DR   GO; GO:0000074; P:regulation of cell cycle; IGI.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001766; TF_Fork_head.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00250; Fork_head; 1.
DR   PRINTS; PR00053; FORKHEAD.
DR   ProDom; PD000425; TF_Fork_head; 1.
DR   SMART; SM0339; FH; 1.
DR   SMART; SM00240; FHA; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00657; FORK_HEAD_1; 1.
DR   PROSITE; PS00658; FORK_HEAD_2; 1.
DR   PROSITE; PS50039; FORK_HEAD_3; 1.
KW   DNA-binding; Nuclear protein.
FT   DOMAIN   76   142   FHA.
FT   DNA_BIND   302   393   FORK-HEAD.
SQ   SEQUENCE   484 AA; 53490 MW; 7159073EB979C489 CRC64;

Query Match   16.1%; Score 61; DB 1; Length 484;
Best Local Similarity   24.7%; Pred. No. 21;
Matches   22; Conservative 15; Mismatches 34; Indels 18; Gaps 2;

Qy   3 LRKREWTIGR---------RRGCDLSFPSNKLVSGDHCRIVDEKSGQVTLEDT 47
               ::||||||| :|: |: ||| |::|: :|||:|:|:||
Db   71 VQKLEVTIGRNTDSLNLNAVPGTVVKNIDIDLGPAKIVSRKHAAIRFNLESGSWELQIF 130

Qy   48 STSGTVINKLKVVKKQTCP---LQTGDVI 73
               :||| |:||:|| |: ||:| :|
Db   131 GRNGAKVNFRRIPTGPDSPPTVLQSGCII 159

RESULT 11
ID   CHK2_MOUSE   STANDARD;   PRT;   546 AA.
AC   Q9Z265;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Serine/threonine-protein kinase Chk2 (EC 2.7.1.37).
GN   CHEK2 OR CHK2 OR RAD53.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99055399; PubMed=9836640;
RA   Matsuoka S., Huang M., Elledge S.J.;
RT   "Linkage of ATM to cell cycle regulation by the Chk2 protein kinase.";
RL   Science 282:1893-1897(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NMRI; TISSUE=Mammary gland;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M, Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: CONTROLS CELL CYCLE CHECKPOINT. MAY PARTICIPATE IN
CC   TRANSDUCTION OF THE DNA DAMAGE AND REPLICATIONAL STRESS SIGNALS.
CC   INHIBITS CDC25 PHOSPHATASE BY PHOSPHORYLATING IT, PREVENTING THE
CC   ENTRY INTO MITOSIS. MAY HAVE A ROLE IN MEIOSIS AS WELL.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- ENZYME REGULATION: KINASE ACTIVITY IS UPREGULATED BY
CC   AUTOPHOSPHORYLATION. RAPIDLY PHOSPHORYLATED IN RESPONSE TO DNA
CC   DAMAGE AND TO REPLICATION BLOCK (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. CDS1
CC   subfamily.
CC   -!- SIMILARITY: Contains 1 FHA domain.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AF086905; AAC83694.1; -.
DR   EMBL; BC056617; AAH56617.1; -.
DR   HSSP; Q63450; 1AO6.
MGD; MGI:1355321; Chek2.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Cell cycle; Phosphorylation; Nuclear protein.
FT   DOMAIN   117   179   FHA.
FT   DOMAIN   224   490   PROTEIN KINASE.
FT   NP_BIND   230   238   ATP (BY SIMILARITY).
FT   BINDING   253   253   ATP (BY SIMILARITY).
FT   ACT_SITE   351   351   BY SIMILARITY.
SQ   SEQUENCE   546 AA; 61088 MW; A7949FB5572CDAA CRC64;

Query Match   16.1%; Score 61; DB 1; Length 546;
Best Local Similarity   28.6%; Pred. No. 24;

```
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zharg G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [4]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RX   MEDLINE=2426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM SHORT).
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=2426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC   -!- FUNCTION: May have a role in germ line establishment.
CC   -!- SUBCELLULAR LOCATION: Nuclear; speckled subnuclear compartment.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=O61267-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=O61267-2; Sequence=VSP_004865;
CC   -!- TISSUE SPECIFICITY: In stage 3 embryos, both isoforms are
CC       expressed in both somatic and pole cell nuclei. Expression in pole
CC       cell nuclei is sustained until stage 9 and weakly expressed after
CC       pole cell invagination into the abdominal cavity.
CC   -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically in
CC       adult females. Levels of the long isoform remain fairly constant
CC       from ovaries to embryos, the levels of short isoform decrease
CC       dramatically.
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. CDS1
CC       subfamily.
CC   -!- SIMILARITY: Contains 1 FHA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on  its
CC   use  by non-profit institutions as long  as its content is in  no  way
CC   modified and this statement is not removed. Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB007821; BAA28755.1; -.
DR   EMBL; AB007822; BAA28756.1; -.
DR   EMBL; U87984; AAB49642.1; -.
DR   EMBL; AE003665; AAF53867.2; -.
DR   EMBL; AE003665; AAN11062.1; -.
DR   EMBL; AY070549; AAL48020.1; -.
DR   HSSP; Q63450; 1A06.
DR   FlyBase; FBgn0019686; lok.
DR   GO; GO:0005634; C:nucleus; IEP.
```

```
DR   GO; GO:0004674; F:protein serine/threonine kinase activity; NAS.
DR   GO; GO:0008630; P:DNA damage response, signal transduction re. . .; IMP.
DR   GO; GO:0000077; P:DNA damage response, signal transduction re. . .; IMP.
DR   GO; GO:0007281; P:germ-cell development; IEP.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; NAS.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00006; FHA_DOMAIN; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Nuclear protein; Alternative splicing.
FT   DOMAIN        69    129       FHA.
FT   DOMAIN       174    441       PROTEIN KINASE.
FT   NP_BIND      180    188       ATP (BY SIMILARITY).
FT   BINDING      203    203       ATP (BY SIMILARITY).
FT   ACT_SITE     303    303       BY SIMILARITY.
FT   VARSPLIC      46     62       Missing (in isoform Short).
FT                                 /FTId=VSP_004865.
SQ   SEQUENCE   476 AA;  54261 MW;  58D583E015C4E626 CRC64;

Query Match             16.3%;  Score 61.5;  DB 1;  Length 476;
Best Local Similarity   25.9%;  Pred. No. 18;
Matches    22;  Conservative  11;  Mismatches  33;  Indels  19;  Gaps   2;

Qy     3  LRKREWTIGRRRGCDLSFPSNKL-------------VSGDHCRIVVDEKSGQVTLEDTS  48
             :||:                             |:|::| |: ::|:|
Db    64  LNNDEFTAGRGEANDLILTLNDLPEKILTRISKVHFIIKRANC-----ELTNPVVIQDLS 118

Qy    49  TSGTVINKLKVVKKQTCPLQTGDVI  73
          :||:||:|: : |||  |:|
Db   119  RNGTFVNNEKIGTNRMRILKNDDVI 143

RESULT 10
FKH1_YEAST
ID   FKH1_YEAST   STANDARD;   PRT;   484 AA.
AC   P40466;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Fork head protein homolog 1.
GN   FKH1 OR YIL131C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RA   Zhu G., Davis T.N.;
RL   Submitted (XXX-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   STRAIN=S288c / AB972;
RX   MEDLINE=97313266; PubMed=9169870;
RA   Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA   Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA   Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
RA   Moule S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA   Skelton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
RL   Nature 387:84-87(1997).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
```

```
SQ   SEQUENCE   445 AA;   51151 MW;   6D62D79E9A1B45B0 CRC64;

     Query Match         16.8%;   Score 63.5;   DB 1;   Length 445;
     Best Local Similarity   27.8%;   Pred. No. 10;
     Matches   20;   Conservative   14;   Mismatches   27;   Indels   11;   Gaps   3;

Qy     9 TIGRRGCD--LSFPSNKLVSGDHCRI----VVDEKSGQVTLEDTSTSGTVINKLKVVK 61
         :: |||     |:||:|   | |:     |   :|||:| ||:|::|::| |::| | :
Db    63 SVGRSNTCNYQLLQFTA----SYKHFRVYSVLIDDDMDPLVYCEDQSSNGTFLNHRLIGK 118

Qy    62 KQTCPIQTGDVI 73
         :: :|:|| |:|
Db   119 GNSVLLSDGDIL 130


RESULT 9
LOK_DROME
ID   LOK_DROME        STANDARD;      PRT;   476 AA.
AC   061267; 061268; P91876; Q8SZS3;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ovarian-specific serine/threonine-protein kinase Lok (EC 2.7.1.-)
DE   (Loki protein) (dMNK).
GN   LOK OR CG10895.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT), FUNCTION, SUBCELLULAR
RP   LOCATION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC   STRAIN=Canton-S; TISSUE=Embryo;
RX   MEDLINE=98175876; PubMed=9507063;
RA   Oishi I., Sugiyama S., Otani H., Yamamura H., Nishida Y., Minami Y.;
RT   "A novel Drosophila nuclear protein serine/threonine kinase expressed
RT   in the germline during its establishment.";
RL   Mech. Dev. 71:49-63(1998).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM SHORT).
RC   STRAIN=Oregon-R; TISSUE=Ovary;
RA   Larochelle S., Suter B.;
RT   "Identification of a novel ovarian specific protein kinase.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew-I., Dietz S.M.,
RA   Dodson M.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraro C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
```

```
RA   Lyne M.H., Bryant J.A., Aves S.J.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Probable protein kinase required for meiotic
CC       recombination.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC   -!- SIMILARITY: Contains 1 FHA domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z71478; CAA96101.1; -.
DR   EMBL; Z98596; CAB11196.1; -.
DR   PIR; T43420; T43420.
DR   HSSP; P00518; 1PHK.
DR   GeneDB_SPombe; SPAC14C4.03; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_Thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; Meiosis.
FT   DOMAIN       62    116       FHA.
FT   DOMAIN      160    421       PROTEIN KINASE.
FT   NP_BIND     166    174       ATP (BY SIMILARITY).
FT   ACT_SITE    281    281       BY SIMILARITY.
FT   BINDING     189    189       ATP (BY SIMILARITY).
```

```
                                        VSP_006010;
       Name=3;
          IsoId=P97836-3; Sequence=VSP_006007, VSP_006008;
   -!- TISSUE SPECIFICITY: Expressed in brain and testis.
   -!- SIMILARITY: BELONGS TO THE SAPAP FAMILY.
   --------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
   --------------------------------------------------------------
DR   EMBL; U67987; AAC53054.1; -.
DR   EMBL; AB003594; BAA24265.1; -.
DR   EMBL; U67137; AAB48587.1; -.
DR   InterPro; IPR005026; GKAP.
DR   Pfam; PF03359; GKAP; 1.
KW   Membrane; Alternative splicing.
FT   VARSPLIC      1    298       Missing (in isoform 2 and isoform 3).
FT                                /FTId=VSP_006007.
FT   VARSPLIC    299    324       QKASVNMDQAVVKSEACQQERSCQYL -> MIDLFKAEWVS
FT                                SVCVQVSRNGRTD (in isoform 2 and isoform
FT                                3).
FT                                /FTId=VSP_006008.
FT   VARSPLIC    325    325       Q -> QVW (in isoform 2).
FT                                /FTId=VSP_006009.
FT   VARSPLIC    547    574       Missing (in isoform 2).
FT                                /FTId=VSP_006010.
FT   MUTAGEN     990    990       T->A: ABOLISHES INTERACTION WITH SHANK1.
FT   MUTAGEN     992    992       L->A: ABOLISHES INTERACTION WITH SHANK1.
FT   CONFLICT    636    636       S -> T (IN REF. 1 AND 2).
SQ   SEQUENCE    992 AA;  110177 MW;  44BAF99BC0C14C099 CRC64;

Query Match            17.5%;  Score 66;  DB 1;  Length 992;
Best Local Similarity  26.9%;  Pred. No. 13;
Matches     21;  Conservative   11;  Mismatches   30;  Indels   16;  Gaps   1;

Qy       7 EWTIGRRRGCDLSFPSNKLVSGDHCRIVDEKSG------------QVTLEDTSTS 50
           |||         || ||  :  |::  |||:||               :| ||  ||
Db     330 EWTGYTPRGKDDEIPCRRMRSGSYIKAMGDBDSGGSDTSPKPSPKVAARRESYIKATQPS 389

Qy      51 GTVINKLKVVKQTCPLQ 68
           ||| |  :|:  :: ||
Db     390 LTELTTLKISNEHSPKLQ 407

RESULT 7
ID   VGLM_BUNSH    STANDARD;       PRT;   1441 AA.
AC   P04875;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   M polyprotein precursor [Contains: Glycoprotein G2; Nonstructural
DE   protein NS-M; Glycoprotein G1].
GN   M.
OS   Bunyavirus snowshoe hare.
OC   Viruses; ssRNA negative-strand viruses; Bunyaviridae; Orthobunyavirus.
OX   NCBI_TaxID=11580;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85018241; PubMed=6091326;
RA   Eshita Y., Bishop D.H.L.;
RT   "The complete sequence of the M RNA of snowshoe hare bunyavirus
RT   reveals the presence of internal hydrophobic domains in the viral
RT   glycoprotein.";
RL   Virology 137:227-240(1984).
RN   [2]
RP   SEQUENCE OF 1-46 FROM N.A.
RX   MEDLINE=82216937; PubMed=7086954;
RA   Clerx-Van Haaster C.M., Akashi H., Auperin D.D., Bishop D.H.L.;
```

```
RT   "Nucleotide sequence analyses and predicted coding of bunyavirus
RT   genome RNA species.";
RL   J. Virol. 41:119-128(1982).
RN   [3]
RP   PROCESSING.
RX   MEDLINE=89073745; PubMed-2974218;
RA   Fazakerley J.K., Gonzalez-Scarano F., Strickler J., Dietz-Schold B.,
RA   Karush F., Nathanson N.;
RT   "Organization of the middle RNA segment of snowshoe hare Bunyavirus.";
RL   Virology 167:422-432(1988).
   -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins
CC       including nonstructural protein NS-M, glycoprotein G1, and
CC       glycoprotein G2.
   -!- SIMILARITY: Belongs to the bunyaviruses M polyprotein family.
   --------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
   --------------------------------------------------------------
DR   EMBL; K02539; AAA47827.1; -.
DR   EMBL; J02392; -; NOT_ANNOTATED_CDS.
DR   PIR; A04102; GNVUSV.
DR   InterPro; IPR005167; Bunya_G1.
DR   InterPro; IPR005168; Bunya_G2.
DR   Pfam; PF03557; Bunya_G1; 1.
DR   Pfam; PF03563; Bunya_G2; 1.
KW   Polyprotein; Glycoprotein; Transmembrane; Nonstructural protein;
KW   Signal.
FT   SIGNAL        1     13
FT   CHAIN        14   1441       M POLYPROTEIN.
FT   CHAIN        14    299       GLYCOPROTEIN G2.
FT   CHAIN       300    473       NONSTRUCTURAL PROTEIN NS-M.
FT   CHAIN       474   1441       GLYCOPROTEIN G1.
FT   CARBOHYD     57     57       N-LINKED (GLCNAC...)      (POTENTIAL).
FT   CARBOHYD    245    245       N-LINKED (GLCNAC...)      (POTENTIAL).
FT   CARBOHYD    490    490       N-LINKED (GLCNAC...)      (POTENTIAL).
FT   CARBOHYD   1177   1177       N-LINKED (GLCNAC...)      (POTENTIAL).
SQ   SEQUENCE   1441 AA;  162389 MW;  51F01DB268D1A0BB CRC64;

Query Match            16.9%;   Score 64;  DB 1;  Length 1441;
Best Local Similarity  22.7%;   Pred. No. 32;
Matches     20;  Conservative   13;  Mismatches   27;  Indels   28;  Gaps   3;

Qy       6 REWTIGRRRGC-----------------DLSFPSNKLVSGDHCRIVDEKSGQVTLEDTS 48
           |:|:                         |||  |:::|:||||
Db      69 RKWLVSDWHDCRPKKITGGHINVIEVGDDLSLHTESVVCSADCTIGVDKETAQVRLQTDT 128

Qy      49 T-----SGTVI-----NKLKVVKKQTC 65
           |:||    |::|      |::|::
Db     129 TNHFEIAGTIVKSGWFKSTTYITLDQTC 156

RESULT 8
ID   MEK1_SCHPO    STANDARD;       PRT;    445 AA.
AC   Q10292;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Meiosis-specific serine/threonine-protein kinase mek1 (EC 2.7.1.37).
GN   MEK1 OR SPAC14C4.03.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
```

```
CC   -!- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1; Synonyms=DAP1-alpha;
CC         IsoId=O14490-1; Sequence=Displayed;
CC       Name=2; Synonyms=DAP1-beta;
CC         IsoId=O14490-2; Sequence=VSP_006003, VSP_006004, VSP_006005,
CC                         VSP_006006;
CC       Name=3;
CC         IsoId=O14490-3; Sequence=VSP_006003, VSP_006004, VSP_006005,
CC                         VSP_006006;
CC   -!- TISSUE SPECIFICITY: Expressed in brain.
CC   -!- SIMILARITY: BELONGS TO THE SAPAP FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB000277; BAA23258.1; -.
DR   EMBL; AB000276; BAA23257.1; -.
DR   EMBL; U67988; AAC51119.1; -.
DR   PIR; T00014; T00014.
DR   Genew; HGNC:2905; DLGAP1.
DR   MIM; 605445; -.
DR   GO; GO:0007273; P:regulation of synapse; NAS.
DR   InterPro; IPR005026; GKAP.
DR   Pfam; PF03359; GKAP; 1.
KW   Membrane; Alternative splicing.
FT   DOMAIN        622    630       POLY-THR.
FT   VARSPLIC        1    302       Missing (in isoform 2 and isoform 3).
FT                                  /FTId=VSP_006003.
FT   VARSPLIC      303    319       MVKSESCQQERSCQYLQ -> MNLIFHKDILFGIPANK
FT                                  (in isoform 2 and isoform 3).
FT                                  /FTId=VSP_006004.
FT   VARSPLIC      909    929       ERRAPPPVPKKPAKGPAPLIR -> VEQCRFCMVHLKTCTN
FT                                  TGQSK (in isoform 3).
FT                                  /FTId=VSP_006005.
FT   VARSPLIC      930    977       Missing (in isoform 3).
FT                                  /FTId=VSP_006006.
FT   CONFLICT      734    734       A -> P (IN REF. 2).
FT   CONFLICT      740    740       S -> T (IN REF. 2).
FT   CONFLICT      752    753       AA -> SP (IN REF. 2).
SQ   SEQUENCE      977 AA;  108872 MW;  670F72B17D9BE667 CRC64;

Query Match             17.5%;   Score 66;  DB 1;  Length 977;
Best Local Similarity   26.9%;  Pred. No. 12;
Matches   21;  Conservative  11;  Mismatches  30;  Indels  16;  Gaps  1;

Qy     7  EWTIGRRRGCDLSFPSNKLVSGDHCRIVVDBKSG------------QVTLEDTSTS 50
          ||||    ||    : |::|::|  ||                  :  :||
Db   324  EWTGYTPRGKDDEIPCRRMRSGSYIKAMGDEDSGDSDTSPKPSPSKVAARRSYLKATQPS 383

Qy    51  GTVINKLKVVKKQTCPIQ 68
          |:||::|||::: :||
Db   384  LTELTTLKISNEHSPKLQ 401

RESULT 6
DLP1_RAT
ID   DLP1_RAT    STANDARD;    PRT;   992 AA.
AC   P97836; O54773; P97841;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Disks large-associated protein 1 (DAP-1) (Guanylate kinase-associated
DE   protein) (rGKAP) (SAP90/PSD-95-associated protein 1) (SAPAP1) (PSD-
DE   95/SAP90 binding protein 1).
GN   DLGAP1 OR GKAP.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=97177144; PubMed=9024696;
RA   Kim E., Naisbitt S., Hsueh Y.-P., Rao A., Rothschild A., Craig A.M.,
RA   Sheng M.;
RT   "GKAP, a novel synaptic protein that interacts with the guanylate
RT   kinase-like domain of the PSD-95/SAP90 family of channel clustering
RT   molecules.";
RL   J. Cell Biol. 136:669-678(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RX   MEDLINE=98089008; PubMed=9428732;
RA   Kawashima N., Takamiya K., Sun J., Kitabatake A., Sobue K.;
RT   "Differential expression of isoforms of PSD-95 binding protein
RT   (GKAP/SAPAP1) during rat brain development.";
RL   FEBS Lett. 418:301-304(1997).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RX   MEDLINE=97277335; PubMed=9115257;
RA   Takeuchi M., Hata Y., Hirao K., Toyoda A., Irie M., Takai Y.;
RT   "SAPAPs. A family of PSD-95/SAP90-associated proteins localized at
RT   postsynaptic density.";
RL   J. Biol. Chem. 272:11943-11951(1997).
RN   [4]
RP   INTERACTION WITH DLG4 AND SHANK PROTEINS.
RX   MEDLINE=99458653; PubMed=10527873;
RA   Boeckers T.M., Winter C., Smalla K.-H., Kreutz M.R., Bockmann J.,
RA   Seidenbecher C., Garner C.C., Gundelfinger E.D.;
RT   "Proline-rich synapse-associated proteins ProSAP1 and ProSAP2 interact
RT   with synaptic proteins of the SAPAP/GKAP family.";
RL   Biochem. Biophys. Res. Commun. 264:247-252(1999).
RN   [5]
RP   INTERACTION WITH DLG4 AND SHANK1.
RC   TISSUE=Brain;
RX   MEDLINE=99419021; PubMed=10488079;
RA   Yao I., Hata Y., Hirao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;
RT   "Synamon, a novel neuronal protein interacting with synapse-associated
RT   protein 90/postsynaptic density-95-associated protein.";
RL   J. Biol. Chem. 274:27463-27466(1999).
RN   [6]
RP   INTERACTION WITH SHANK PROTEINS, AND MUTAGENESIS ALA-990 AND ALA-992.
RX   MEDLINE=99360650; PubMed=10433268;
RA   Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtschanoff J.,
RA   Weinberg R.J., Worley P.F., Sheng M.;
RT   "Shank, a novel family of postsynaptic density proteins that binds to
RT   the NMDA receptor/PSD-95/GKAP complex and cortactin.";
RL   Neuron 23:569-582(1999).
RN   [7]
RP   INTERACTION WITH AIP1.
RX   MEDLINE=98361985; PubMed=9694864;
RA   Hirao K., Hata Y., Ide N., Takeuchi M., Irie M., Yao I., Deguchi M.,
RA   Toyoda A., Sudhof T.C., Takai Y.;
RT   "A novel multiple PDZ domain-containing molecule interacting with
RT   N-methyl-d-aspartate receptors and neuronal cell adhesion proteins.";
RL   J. Biol. Chem. 273:21105-21110(1998).
CC   -!- FUNCTION: May play a role in the molecular organization of
CC       synapses and neuronal cell signaling. Could be an adapter protein
CC       linking ion channel to the subsynaptic cytoskeleton. May induce
CC       enrichment of PSD-95/SAP90 at the plasma membrane.
CC   -!- SUBUNIT: Interacts with DLG1, DLG4/PSD-95, APC and AIP1. The C-
CC       terminus binds the PD2 domain of the SHANK1, SHANK2 and SHANK3. Is
CC       part of a complex with DLG4/PSD-95 and SHANK1, SHANK2 or SHANK3.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=1;
CC         IsoId=P97836-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P97836-2; Sequence=VSP_006007, VSP_006008, VSP_006009,
```

DR   HSSP; Q63450; 1A06.
DR   GermOnline; 140343; -.
DR   SGD; S0002259; DUN1.
DR   GO; GO:0004672; F:protein kinase activity; IDA.
DR   GO; GO:0000077; P:DNA damage response, signal transduction re. . .; IMP.
DR   GO; GO:0006468; P:protein amino acid phosphorylation; IDA.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_Thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Nuclear protein; Phosphorylation; DNA damage.
FT   DOMAIN   56   112   FHA.
FT   DOMAIN   200   480   PROTEIN KINASE.
FT   NP_BIND   206   214   ATP (BY SIMILARITY).
FT   BINDING   229   229   ATP (BY SIMILARITY).
FT   ACT_SITE   328   328   BY SIMILARITY.
FT   MUTAGEN   229   229   K->R: LOSS OF KINASE FUNCTION.
FT   MUTAGEN   328   328   D->A: LOSS OF KINASE FUNCTION.
SQ   SEQUENCE   513 AA; 58632 MW; 4990F24F024702D7 CRC64;

Query Match   19.0%; Score 72; DB 1.3; Length 513;
Best Local Similarity 35.7%; Pred. No. 1.3;
Matches   25; Conservative   10; Mismatches   29; Indels   6; Gaps   3;

Qy   9 TIGRRRGCD--LSFPSNKLVSGDHCRIVVDEKSGQ---VTLEDTSGTVINKLKVVKKQ 63
         |||| || ||| :   : :| :|| | | :| ||||||||||:| ||
Db   57 TIGRSRSCDVILSEPDISTFHAEFHLLQMDVDNFQRNLINVIDKSRNGTFINGNRLVKKD 116

Qy   64 TCPLQTGDVI 73
          : |||
Db 117 YI-LKNGDRI 125

RESULT 4
PGK_MYCPN
ID   PGK_MYCPN   STANDARD;   PRT;   409 AA.
AC   P78018;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phosphoglycerate kinase (EC 2.7.2.3).
GN   PGK OR MPN429 OR MP412.
OS   Mycoplasma pneumoniae.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2104;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 29342 / M129;
RX   MEDLINE=97105885; PubMed=8948633;
RA   Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA   Herrmann R.;
RT   "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT   pneumoniae.";
RL   Nucleic Acids Res. 24:4420-4449(1996).
CC   -!- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-
CC   phospho-D-glyceroyl phosphate.
CC   -!- PATHWAY: Second phase of glycolysis; second step.
CC   -!- SUBUNIT: Monomer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the phosphoglycerate kinase family.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration

CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AE000040; AAB96060.1; -.
DR   PIR; S73738; S73738.
DR   HSSP; P36204; 1VPE.
DR   HAMAP; MF_00145; -; 1.
DR   InterPro; IPR001576; PGK.
DR   Pfam; PF00162; PGK; 1.
DR   PRINTS; PR00477; PHGLYCKINASE.
DR   PROSITE; PS00111; PGLYCERATE_KINASE; 1.
KW   Transferase; Kinase; Glycolysis; Complete proteome.
SQ   SEQUENCE   409 AA; 44212 MW; 6AF230188D398731 CRC64;

Query Match   18.8%; Score 71; DB 1; Length 409;
Best Local Similarity 29.2%; Pred. No. 1.3;
Matches   19; Conservative   14; Mismatches   20; Indels   12; Gaps   2;

Qy   19 SFPSNKLVSGDHCRIV---------------VDEKSGQVTLEDTSGTVINKLKVVKKQTCPLQ 68
          :| :| : |||| :| :|| | ||||||||||:| ||| |||||
Db   41 SLDIKYLLGHNCKIVLLSHLSRVKSLDDKKGKKSLQPVASA--LQNLLKNTKVHFCPEN 98

Qy   69 TGDVI 73
         ||| |
Db   99 TGDKV 103

RESULT 5
DLP1_HUMAN
ID   DLP1_HUMAN   STANDARD;   PRT;   977 AA.
AC   O14490; O14489; P78335;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Disks large-associated protein 1 (DAP-1) (Guanylate kinase-associated
DE   protein) (hGKAP) (SAP90/PSD-95-associated protein 1) (SAPAP1) (PSD-
DE   95/SAP90 binding protein 1).
GN   DLGAP1 OR GKAP OR DAP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Brain;
RX   MEDLINE=97431353; PubMed=9286858;
RA   Satoh K., Yanai H., Senda T., Kohu K., Nakamura T., Okumura N.,
RA   Matsumine A., Kobayashi S., Toyoshima K., Akiyama T.;
RT   "DAP-1, a novel protein that interacts with the guanylate kinase-like
RT   domains of hDLG and PSD-95.";
RL   Genes Cells 2:415-424(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RX   MEDLINE=97177144; PubMed=9024696;
RA   Kim E., Naisbitt S., Hsueh Y.-P., Rao A., Rothschild A., Craig A.M.,
RA   Sheng M.;
RT   "GKAP, a novel synaptic protein that interacts with the guanylate
RT   kinase-like domain of the PSD-95/SAP90 family of channel clustering
RT   molecules.";
RL   J. Cell Biol. 136:669-678(1997).
CC   -!- FUNCTION: May play a role in the molecular organization of
CC   synapses and neuronal cell signaling. Could be an adapter protein
CC   linking ion channel to the subsynaptic cytoskeleton. May induce
CC   enrichment of PSD-95/SAP90 at the plasma membrane.
CC   -!- SUBUNIT: Interacts with DLG1, DLG4/PSD-95, APC and AIP1 (By
CC   similarity). Isoform 1 and isoform 2 C-terminus bind the PDZ
CC   domain of SHANK1, SHANK2 and SHANK3 (By similarity). Is part of a
CC   complex with DLG4/PSD-95 and SHANK1, SHANK2 or SHANK3 (By
CC   similarity).

RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Has a role in the DNA replication-monitoring S/G2
CC   checkpoint system. It is responsible for blocking mitosis in the
CC   S phase. It monitors DNA synthesis by interacting with DNA
CC   polymerase alpha and sends a signal to block the onset of mitosis
CC   while DNA synthesis is in progress.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SUBUNIT: Interacts with rad26.
CC   -!- PTM: Autophosphorylated.
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. CDS1
CC   subfamily.
CC   -!- SIMILARITY: Contains 1 FHA domain.
CC   --
DR   EMBL; X85040; CAA59410.1; -.
DR   EMBL; AJ222869; CAA11019.1; -.
DR   EMBL; AL109736; CAB52158.1; -.
DR   PIR; S58882; S58882.
DR   HSSP; Q00534; 1BI7.
DR   GeneDB_SPombe; SPCC18B5.11c; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_Thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Cell cycle; Phosphorylation.
FT   DOMAIN   60   116   FHA.
FT   DOMAIN   167   433   PROTEIN KINASE.
FT   NP_BIND   173   181   ATP (BY SIMILARITY).
FT   BINDING   196   196   ATP (BY SIMILARITY).
FT   ACT_SITE   294   294   BY SIMILARITY.
FT   CONFLICT   61   61   R -> G (IN REF. 1).
FT   CONFLICT   238   238   F -> C (IN REF. 1).

SQ   SEQUENCE   460 AA;   52014 MW;   4CEB963D3376DB54 CRC64;

Query Match   20.5%;   Score 77.5; DB 1; Length 460;
Best Local Similarity   30.6%;   Pred. No. 0.27;
Matches 22; Conservative 9; Mismatches 34; Indels 7; Gaps 2;

Qy   8 WTIGRRRGCDLSPPSNKLVSGDHCRIV-----VDEKSGQVTLEDTSTSGTVINKLKVVK 61
               | :| | : | | .:. | | | ::||::|: |::|::: || ||:|||
Db   60 WRFGRHKSCEVVL-NGPRVSNFHPEIYQGHRNDSDESENVVFLHDHSSNGTFLNPERLAK 118

Qy   62 KQTCPLQTGDVI 73
               | |:|| |||
Db   119 NSRTILSNGDEI 130

RESULT 3
DUN1_YEAST
ID   DUN1_YEAST   STANDARD;   PRT;   513 AA.
AC   P39009;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DNA damage response protein kinase DUN1 (EC 2.7.1.-).
GN   DUN1 OR YDL101C OR D2370.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9084787; PubMed=8261511;
RA   Zhou Z., Elledge S.J.;
RT   "DUN1 encodes a protein kinase that controls the DNA damage response
RT   in yeast.";
RL   Cell 75:1119-1127(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   STRAIN=S288C / FY1679;
RX   MEDLINE=97051597; PubMed=8896274;
RA   Saiz J.E., Buitrago M.J., Garcia R., Revuelta J.L., del Rey F.;
RT   "The sequence of a 20.3 kb DNA fragment from the left arm of
RT   Saccharomyces cerevisiae chromosome IV contains the KIN28, MSS2,
RT   PHO2, POL3 and DUN1 genes, and six new open reading frames.";
RL   Yeast 12:1377-1384(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   STRAIN=S288C / FY1679;
RX   MEDLINE=97082507; PubMed=8923743;
RA   Boskovic J., Saiz J.E., Soler-Mira A., Garcia-Cantalejo J.M.,
RA   Ballesta J.P.G., Saiz J.E., Jimenez A., Remacha M.;
RT   "The sequence of a 16,691 bp segment of Saccharomyces cerevisiae
RT   chromosome IV identifies the DUN1, PWT1, PWT5, SRP14 and DPR1 genes,
RT   and five new open reading frames.";
RL   Yeast 12:1377-1384(1996).
CC   -!- FUNCTION: Transducer of the DNA damage signal.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- PTM: Autophosphorylation increases in response to DNA damage.
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. CDS1
CC   subfamily.
CC   -!- SIMILARITY: Contains 1 FHA domain.
CC   --
DR   EMBL; L25548; AAA16324.1; -.
DR   EMBL; X95644; CAA64912.1; -.
DR   EMBL; Z74149; CAA98668.1; -.
DR   PIR; S43941; S43941.

```
RT   damage-induced transcription in yeast.";
RL   Genes Dev. 8:2401-2415(1994).
RN   [5]
RP   FUNCTION.
RX   MEDLINE=20018334; PubMed=10550056;
RA   Sanchez Y., Bachant J., Wang H., Hu F., Liu D., Tetzlaff M.,
RA   Elledge S.J.;
RT   "Control of the DNA damage checkpoint by chk1 and rad53 protein
RT   kinases through distinct mechanisms.";
RL   Science 286:1166-1171(1999).
RN   [6]
RP   PHOSPHORYLATION.
RX   MEDLINE=20031667; PubMed=10562558;
RA   Pellicioli A., Lucca C., Liberi G., Marini F., Lopes M., Plevani P.,
RA   Romano A., Di Fiore P.P., Foiani M.;
RT   "Activation of Rad53 kinase in response to DNA damage and its effect
RT   in modulating phosphorylation of the lagging strand DNA polymerase.";
RL   EMBO J. 18:6561-6572(1999).
RN   [7]
RP   STRUCTURE BY NMR OF 573-730.
RX   MEDLINE=20057864; PubMed=10588905;
RA   Liao H., Byeon I.-J.L., Tsai M.-D.;
RT   "Structure and function of a new phosphopeptide-binding domain
RT   containing the FHA2 of Rad53.";
RL   J. Mol. Biol. 294:1041-1049(1999).
CC   -!- FUNCTION: Controls S-phase checkpoint as well as G1 and G2 DNA
CC       damage checkpoints. Phosphorylates proteins on serine, threonine,
CC       and tyrosine. Prevents entry into anaphase and mitotic exit after
CC       DNA damage via regulation of the Polo kinase CDC5.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- PTM: Autophosphorylated.
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC   -!- SIMILARITY: Contains 2 FHA domains.
```
```
DR   EMBL; M55623; AAA35070.1; -.
DR   EMBL; X96770; CAA65568.1; -.
DR   EMBL; Z73509; CAA97858.1; -.
DR   PIR; A39616; A39616.
DR   PDB; 1DMZ; 06-JAN-00.
DR   PDB; 1FHQ; 18-OCT-00.
DR   PDB; 1FHR; 18-OCT-00.
DR   PDB; 1G3G; 10-JAN-01.
DR   PDB; 1G6G; 13-DEC-00.
DR   PDB; 1J4K; 05-DEC-01.
DR   PDB; 1J4L; 05-DEC-01.
DR   PDB; 1J4O; 05-DEC-01.
DR   PDB; 1J4P; 05-DEC-01.
DR   PDB; 1J4Q; 05-DEC-01.
DR   PDB; 1K2M; 05-DEC-01.
DR   PDB; 1K2N; 05-DEC-01.
DR   PDB; 1K3J; 05-DEC-01.
DR   PDB; 1K3N; 05-DEC-01.
DR   PDB; 1K3Q; 05-DEC-01.
DR   PDB; 1QU5; 17-JAN-01.
DR   GermOnline; 144135; -.
DR   SGD; S0006074; RAD53.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0004712; F:protein threonine/tyrosine kinase activity; IDA.
DR   GO; GO:0004713; F:DNA repair; IMP.
DR   GO; GO:0006119; P:nucleotide, nucleoside, nucleotide and nucl. ..; IGI.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_Thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PP00498; FHA; 2.
DR   Pfam; PP00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00240; FHA; 2.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 2.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding;
KW   Tyrosine-protein kinase; Nuclear protein; Repeat; DNA damage;
KW   Cell cycle; Phosphorylation; 3d-structure.
FT   DOMAIN       66    116   FHA 1.
FT   DOMAIN      198    466   PROTEIN KINASE.
FT   DOMAIN      601    664   FHA 2.
FT   NP_BIND     204    212   ATP (BY SIMILARITY).
FT   BINDING     227    227   ATP (BY SIMILARITY).
FT   ACT_SITE    319    319   BY SIMILARITY.
SQ   SEQUENCE   821 AA;  91962 MW;  84A9612229CA72D1 CRC64;

Query Match            28.4%;   Score 107.5;  DB 1;  Length 821;
Best Local Similarity  37.7%;   Pred. No. 0.0002;
Matches   26;  Conservative   9;  Mismatches  33;  Indels   1;  Gaps   1;

Qy      5  KREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQT   64
           : |||||||||||   .::|.:: :: ||||::.: :|:  |  : |:|::: : ||
Db     63  KKVWTFGRNPACDYHLGNISRLSNKHFQILLGB-DGNLLLNDISTNGTWLNGQKVEKNSN  121

Qy     65  CPLQTGDVI   73
           |.|:|:|
Db    122  QLLSQGDEI  130

RESULT 2
CDS1_SCHPO
ID   CDS1_SCHPO     STANDARD;      PRT;    460 AA.
AC   Q09170; O42642;
DT   01-NOV-1995 (Rel. 32, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Protein kinase cds1 (EC 2.7.1.37) (Checkpoint kinase cds1).
GN   CDS1 OR SPCC18B5.11C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A., AND FUNCTION.
RX   MEDLINE=95240713; PubMed=7723827;
RA   Murakami H., Okayama H.;
RT   "A kinase from fission yeast responsible for blocking mitosis in s
RT   phase.";
RL   Nature 374:817-819(1995).
RN   [2]
RP   SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND PHOSPHORYLATION.
RC   STRAIN=972;
RX   MEDLINE=98119835; PubMed=9450932;
RA   Lindsay H.D., Griffiths D.J.F., Edwards R.J., Christensen P.U.,
RA   Murray J.M., Osman F., Walworth N., Carr A.M.;
RT   "S-phase-specific activation of Cds1 kinase defines a subpathway of
RT   the checkpoint response in Schizosaccharomyces pombe.";
RL   Genes Dev. 12:382-395(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 14:37:32 ; Search time 2.46674 Seconds
(without alignments)
1540.951 Million cell updates/sec

Title: US-10-048-046-2_COPY_31_103
Perfect score: 378
Sequence: 1 VLLRKREWTIGRRRGCDLSF..........INKLKVVKKQTCPLQTGDVI 73

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	107.5	28.4	821	1	SPK1_YEAST	P22216	saccharomyc
2	77.5	20.5	460	1	CDS1_SCHPO	Q09170	schizosacch
3	72	19.0	513	1	DUN1_YEAST	P39009	saccharomyc
4	71	18.8	409	1	PGK_MYCPN	P78018	mycoplasma
5	66	17.5	977	1	DLPI_HUMAN	O14490	homo sapien
6	66	17.5	992	1	DLPI_RAT	P97836	rattus norv
7	64	16.9	1441	1	VGLM_BUNSH	P04875	bunyavirus
8	63.5	16.8	445	1	MEK1_SCHPO	Q10292	schizosacch
9	61.5	16.3	476	1	LOK_DROME	O61267	drosophila
10	61	16.1	484	1	FKH1_YEAST	P40466	saccharomyc
11	61	16.1	546	1	CHK2_MOUSE	Q9Z265	mus musculu
12	61	16.1	1433	1	VGLM_BUNYW	P04505	bunyamwera
13	61	16.1	1441	1	VGLM_BUNL7	P09612	bunyavirus
14	60.5	16.0	365	1	TRMU_CORGL	Q8NR24	corynebacte
15	60	15.9	353	1	Y634_CHLMU	Q9PK39	chlamydia m
16	60	15.9	403	1	GCST_MOUSE	Q8CFA2	mus musculu
17	60	15.9	952	1	YKI5_CAEEL	P46012	caenorhabdi
18	60	15.9	969	1	UVRA_MYCLE	Q9CC24	mycobacteri
19	59.5	15.7	465	1	TM15_HUMAN	Q9C019	homo sapien
20	59.5	15.7	485	1	RNF8_HUMAN	O76064	homo sapien
21	59.5	15.7	809	1	DCML_OLICA	P19919	oligotropha
22	59	15.6	592	1	YTRP_PSEPU	P40604	pseudomonas
23	59	15.6	1045	1	HMD2_YEAST	P12684	saccharomyc
24	58.5	15.5	181	1	KAD_MYCTU		mycobacteri
25	58.5	15.5	209	1	GPNA_RHIET	Q8KL44	rhizobium e
26	58.5	15.5	523	1	P60_LISSE	Q01838	listeria se
27	58.5	15.5	814	1	SLA1_BACAA	P49051	bacillus an
28	58	15.3	497	1	MEK1_YEAST	P24719	saccharomyc
29	58	15.3	522	1	YNL6_YEAST	P53922	saccharomyc
30	58	15.3	524	1	HUNB_TRICA	Q01791	tribolium c
31	58	15.3	746	1	TAGF_BACSU	P13485	bacillus su
32	58	15.3	888	1	SYA_ZYMMO	Q9mn8	zymomonas m
33	58	15.3	891	1	VP4A_VACCC	P20642	vaccinia vi
34	58	15.3	891	1	VP4A_VACCV	P16715	vaccinia vi
35	58	15.3	892	1	VP4A_VARV	P33817	variola vir
36	58	15.3	1009	1	WS14_CAEEL	P41846	caenorhabdi
37	57.5	15.2	469	1	YOJ8_CAEEL	P34631	caenorhabdi
38	57.5	15.2	589	1	SYFB_HUMAN	Q9nsd9	homo sapien
39	57.5	15.2	780	1	TREB_YEAST	P35172	saccharomyc
40	57	15.1	475	1	FA10_CHICK	P25155	gallus gall
41	57	15.1	936	1	FHL1_YEAST	P39521	saccharomyc
42	57	15.1	1333	1	ADO_RAT	Q9zou5	rattus norv
43	56.5	14.9	215	1	KAD_METAC	Q8trs3	methanosarc
44	56.5	14.9	334	1	YMOI_RHIME	P49305	rhizobium m
45	56.5	14.9	376	1	DNAJ_DEIRA	Q9rug2	deinococcus

ALIGNMENTS

RESULT 1
SPK1_YEAST
ID   SPK1_YEAST   STANDARD;   PRT;   821 AA.
AC   P22216;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Protein kinase SPK1 (EC 2.7.1.-) (Serine-protein kinase 1).
GN   SPK1 OR SAD1 OR RAD53 OR MEC2 OR YPL153C OR P2588.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91117267; PubMed=1899289;
RA   Stern D.F., Zheng P., Beidler D.R., Zerillo C.;
RT   "Spk1, a new kinase from Saccharomyces cerevisiae, phosphorylates
RT   proteins on serine, threonine, and tyrosine.";
RL   Mol. Cell. Biol. 11:987-1001(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97103777; PubMed=8948103;
RA   Purnelle B., Coster F., Goffeau A.;
RT   "The sequence of 55 kb on the left arm of yeast chromosome XVI
RT   identifies a small nuclear RNA, a new putative protein kinase and two
RT   new putative regulators.";
RL   Yeast 12:1483-1492(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313271; PubMed=9169875;
RA   Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,
RA   Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA   Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA   Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA   Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA   Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA   Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA   Hunicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA   Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA   Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA   Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA   Petel F.X., Pohl T.M., Purnelle B., Schafer M., Scharfe M.,
RA   Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA   Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA   Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA   Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL   Nature 387:103-105(1997).
RN   [4]
RP   FUNCTION.
RX   MEDLINE=95047382; PubMed=7958905;
RA   Allen J.B., Zhou Z., Siede W., Friedberg E.C., Elledge S.J.;
RT   "The SAD1/RAD53 protein kinase controls multiple checkpoints and DNA

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 14, 2004, 20:06:30 ; Search time 1483 Seconds
                 (without alignments)
                 7674.257 Million cell updates/sec

Title: US-10-048-046-1
Perfect score: 2679
Sequence: 1 aagaattcggcacgaggccg..........acaaaaaaaaaaaaaaa 2679

Scoring table: OLIGO_NUC
                 Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size :   0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :   N_Geneseq_29Jan04:*
                 1: geneseqn1980s:*
                 2: geneseqn1990s:*
                 3: geneseqn2000s:*
                 4: geneseqn2001as:*
                 5: geneseqn2001bs:*
                 6: geneseqn2002s:*
                 7: geneseqn2003as:*
                 8: geneseqn2003bs:*
                 9: geneseqn2003cs:*
                 10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description		
1	2679	100.0	2679	5	AAF30352	Aaf30352 Human chf	24	139	5.2	825	4	AAK58374	Human imm
2	2133	79.6	2639	4	AAH14542	Aah14542 Human CDN	25	78	2.9	617	6	ABN60913	Human can
3	1995	74.5	1995	5	AAF89709	Aaf89709 Nucleotid	26	77	2.9	476	8	ACH14968	Human adu
4	1618	60.4	3181	6	ABN59646	Abn59646 Novel hum	27	60	2.2	60	6	ABN33472	Human spl
5	1465	54.7	3138	4	AAH14556	Aah14556 Human CDN	28	26	1.0	860	2	AAQ22007	Sequence
6	1024	38.2	2259	4	AAI88903	Aai88903 Human pol	29	25	0.9	1085	3	AAA96229	cDNA enco
7	765	28.6	1311	4	AAS25843	Aas25843 Human CDN	30	25	0.9	1963	6	AAD39143	Human tra
8	765	28.6	1311	7	ABX73184	Abx73184 Human nov	31	25	0.9	1963	9	AAD37442	Human tra
9	633	23.6	2186	7	ADA52592	Ada52592 Human cod	32	25	0.9	2000	6	ABZ15532	Arabidops
10	478	17.8	816	4	AAH06828	Aah06828 Human CDN	33	25	0.9	2000	7	ADA69051	Arabidops
11	468	17.5	824	4	AAH06846	Aah06846 Human CDN	34	25	0.9	95769	7	ADA68659	Arabidops
12	455	17.0	693	4	AAS26304	Aas26304 Human CDN	35	24	0.9	199	6	ABL86644	Human ova
13	455	17.0	693	7	ABX73645	Abx73645 Human nov	36	24	0.9	222	5	AAS29141	cDNA enco
14	426	15.9	518	4	AAH11859	Aah11859 Human CDN	37	24	0.9	222	5	AAS29069	cDNA enco
15	286	10.7	575	3	AAA44336	Aaa44336 Human sec	38	24	0.9	222	6	ABS68281	cDNA enco
16	267	10.0	357	4	AAS26143	Aas26143 Human CDN	39	24	0.9	222	6	ABS68209	cDNA enco
17	267	10.0	357	7	ABX73484	Abx73484 Human nov	40	24	0.9	222	9	ADC25203	Human CDN
18	238	8.9	354	4	AAS26563	Aas26563 Human CDN	41	24	0.9	222	9	ADC25275	Human CDN
19	238	8.9	354	4	AAS26144	Aas26144 Human CDN	42	24	0.9	291	5	ABV20036	Human pro
20	238	8.9	354	7	ABX73904	Abx73904 Human nov	43	24	0.9	413	5	ABV49795	Human pro
21	238	8.9	354	7	ABX73485	Abx73485 Human nov	44	24	0.9	474	8	ACH23460	Human adu
22	217	8.1	449	8	ACH26450	Ach26450 Human adu	45	24	0.9	525	5	ABV60912	Human pro
23	169	6.3	13836	5	ABA16633	Aba16633 Human ner							

ALIGNMENTS

RESULT 1
AAF30352
ID   AAF30352 standard; cDNA; 2679 BP.
XX
AC   AAF30352;
XX
DT   14-MAY-2001 (first entry)
XX
DE   Human chfr cDNA encoding checkpoint with FHA and ring finger protein.
XX
KW   Checkpoint with forkhead associated domain and ring finger; Chfr; human;
KW   mitosis; cell cycle; tumour; diagnosis; antitumour; drug screening;
KW   ubiquitin-protein ligase; ss.
XX
OS   Homo sapiens.
XX
FH   Location/Qualifiers
FT   Key
FT   CDS   91. .2085
FT   /*tag= a
FT   replace(1828,A)
FT   /*tag= b
FT   mutation   /note= "alters codon GTG (Val) to ATG (Met) in chfr from
FT   human cancer cell line U2OS"
XX
PN   WO200109150-A2.
XX
PD   08-FEB-2001.
XX
PF   14-JUN-2000; 2000WO-US016391.
XX
PR   29-JUL-1999; 99US-0146194P.
XX
PA   (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI   Halazonetis T, Scolnick D;
XX
DR   WPI; 2001-182927/18.
DR   P-PSDB; AAB20219.
XX
PT   Novel nucleic acid sequence of mitotic checkpoint gene encoding a
PT   checkpoint with forkhead-associated domain and ring finger protein, for
PT   diagnosing tumorigenic cells and in screening for anticancer drugs.
XX
PS   Claim 2(a); Fig 4A-D; 85pp; English.
XX
CC   The present sequence is that of cDNA encoding the human mitotic

CC   checkpoint protein Chfr (see AAB20219) having a forkhead associated
CC   domain (FHA) and a ring finger domain. The protein is required for
CC   regulation of the transition of cells from prophase to metaphase during
CC   mitosis. It has ubiquitin-protein ligase activity. The Chfr checkpoint
CC   was evident in primary human cells, but was inactivated in 4 of 8 human
CC   cancer cell lines. In U2OS cells, a mutation was identified, which caused
CC   a Val to Met amino acid substitution in the highly conserved C-terminal
CC   Cys-rich region of the Chfr protein. In the absence of the Chfr
CC   checkpoint, cells subjected to mitotic stress condensed their chromosomes
CC   despite failing to separate their chromosomes. Chfr may monitor
CC   centrosome separation. Inactivation of the chfr gene in human cancer is
CC   theorized to underlie the increased sensitivity of cancer cells to
CC   antimitotic drugs. Nucleic acids comprising the present sequence, or
CC   sequences encoding at least amino acids 31-103, 303-346, 476-641 (or
CC   their antisense sequences) are claimed. The chfr cDNA was isolated from
CC   an expressed sequence tag database sequence for cDNAs with FHA motifs.
CC   Claimed methods of determining the tumourigenic potential of a cell
CC   comprise examining the cell for the presence of chfr nucleic acid
CC   (absence or mutation indicating predisposition to tumourigenesis upon
CC   exposure to mitotic stress). Also claimed is a diagnostic agent, e.g. an
CC   antisense fragment of the present sequence, that binds to the chfr
CC   nucleic acid, and a diagnostic kit for detecting tumourigenic potential
CC   of a cell. A composition which inhibits the biological activity of Chfr
CC   may comprise a ligand selected from an antibody or its fragment. The Chfr
CC   inhibitor is used in a claimed method of retarding the growth of a cancer
CC   cells. Chfr polypeptides are useful in screening for drugs which can
CC   inhibit the activity of Chfr in a cancer cell, rendering the cell more
CC   sensitive to additional antitumour therapies
XX
SQ   Sequence 2679 BP; 650 A; 715 C; 764 G; 550 T; 0 U; 0 Other;

Query Match   100.0%;   Score 2679;   DB 5;   Length 2679;
Best Local Similarity   100.0%;   Pred. No.0;
Matches 2679;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 AAGAATTCGGCACGAGGCCGCCAATGTCTTGACAGCGGCGGCGGCGGCCGCCGGCAGCCGGCAGCCGGTTCCGG 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AAGAATTCGGCACGAGGCCGCCAATGTCTTGACAGCGGCGGCGGCGGCCGCCGGCAGCCGGCAGCCGGTTCCGG 60

Qy   61 GTTCGGCGCGGGGGGGGCGGGGATGTGAATCCGGATGCGGAGCGGCCCGGAGGAAGGCAAGCAGTCG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GTTCGGCGCGGGGGGGGCGGGGATGTGAATCCGGATGCGGAGCGGCCCGGAGGAAGGCAAGCAGTCG 120

Qy   121 CCGCCGCCGCAGCCCTGGGGACGGCTCTGCGTCTCGGCGCGGAGACAGAGGTTGCGACCTTTCCTC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCGCCGCCGCAGCCCTGGGGACGGCTCTGCGTCTCGGCGCGGAGACAGAGGTTGCGACCTTTCCTC 180

Qy   181 GTCCTCCTGAGGAAGCGGGAGTGGACCATCGGCGGCGGAGTCCTGAGATTGTAGTGGATGAAAAATCAGGT 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTCCTCCTGAGGAAGCGGGAGTGGACCATCGGCGGCGGAGTCCTGAGATTGTAGTGGATGAAAAATCAGGT 240

Qy   241 CCCAGCAATAAACTGGTCTCTGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CCCAGCAATAAACTGGTCTCTGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGT 300

Qy   301 CAGGTGACACTGAAGATACCAGCACCAGTGAAGATGTGATTAACAGCTGAAGGTTGTT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CAGGTGACACTGAAGATACCAGCACCAGTGAAGATGTGATTAACAGCTGAAGGTTGTT 360

Qy   361 AAGAAGCAGACATGCCCTTTACAGACTGGGAATGTCATCTACTTCGGTGTACAGGAAGAAT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGAAGCAGACATGCCCTTTACAGACTGGGAATGTCATCTACTTCGGTGTACAGGAAGAAT 420

Qy   421 GAACCCGAACAACAACGTGGCATACCTCATGAATCTTAAGTGAAAAGCAAGGCATGACA 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GAACCCGAACAACAACGTGGCATACCTCATGAATCTTAAGTGAAAAGCAAGGCATGACA 480

Qy   481 CAAGAATCCTTTGAAGCTAACAAGGAAAAATGTGTTCCATGGGACCAAAGATACCTCAGGT 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CAAGAATCCTTTGAAGCTAACAAGGAAAAATGTGTTCCATGGGACCAAAGATACCTCAGGT 540

Qy   541 GCAGGTGCAGGGGCGAGGGGCCGATCCCCGGGTCCCTCCGTCGTCGCCCGCCACTCAGGTG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GCAGGTGCAGGGGCGAGGGGCCGATCCCCGGGTCCCTCCGTCGTCGCCCGCCACTCAGGTG 600

Db   541 GCAGGTGCAGGGGCGAGGGGCCGATCCCCGGGTCTCCTCCGTCGTCGCCCGCCACTCAGGTG 600

Qy   601 TGCTTGACGGAACCACAGCCATCAACATCGAGTCAGACCTTCCCACAGCCTCGGCC 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGCTTGACGGAACCACAGCCATCAACATCGAGTCAGACCTTCCCACAGCCTCGGCC 660

Qy   661 TCTTCCACCGAGGCCTTCCTCCTGCAGGGGCGGAGAGCGTTCCTCCAGTGTGGGTCTGGGGGT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TCTTCCACCGAGGCCTTCCTCCTGCAGGGGCGGAGAGCGTTCCTCCAGTGTGGGTCTGGGGGT 720

Qy   721 GGTGGCATCTCCCTAAGGAAGTGGTCCCTGTGGCAAGTGATGAAGTCTCCAGCTTT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GGTGGCATCTCCCTAAGGAAGTGGTCCCTGTGGCAAGTGATGAAGTCTCCAGCTTT 780

Qy   781 GCCTCAGCTCTCCCAGACAGAAAGACTGCGTCCTTTCGTCGTTGGAACCCCAGGATCAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GCCTCAGCTCTCCCAGACAGAAAGACTGCGTCCTTTCGTCGTTGGAACCCCAGGATCAG 840

Qy   841 GAGGATTTGGAGCCCGTGAAGAAGAAATGAGAGGAGATGGGGGACCTTGACCTGAACGGG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 GAGGATTTGGAGCCCGTGAAGAAGAAATGAGAGGAGATGGGGGACCTTGACCTGAACGGG 900

Qy   901 CAGTTGTTGGTCGGCACAACCGGTAGAAATGCCAAACCGTCACGAGGAGTCAGAGCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 CAGTTGTTGGTCGGCACAACCGGTAGAAATGCCAAACCGTCACGAGGAGTCAGAGCA 960

Qy   961 GCGGCTGGGAAGCCAGCACAAGATGGGAGGAGACGCTGACATGCATCATCTGCCAGGACCTG 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GCGGCTGGGAAGCCAGCAGAAGATGGGAGGAGACGCTGACATGCATCATCTGCCAGGACCTG 1020

Qy   1021 CTGCACGACTGCCTGAGTTGCAGCCCTGCATGCAACGTTCTGCGGCGGTTGCTACTCG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CTGCACGACTGCCTGAGTTGCAGCCCTGCATGCAACGTTCTGCGGCGGTTGCTACTCG 1080

Qy   1081 GGTGGATGGAGCGCTCGCCCTGTCCTACCTGCGTCCCGTGCCCGGGAGCGGGATCTGT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GGTGGATGGAGCGCTCGCCCTGTCCTACCTGCGTCCCGTGCCCGGGAGCGGGATCTGT 1140

Qy   1141 AAAAACCACATCCTCAACAACCTCGTGGAAGCATACCTCATCTCAGCATCAGACAAGAGT 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 AAAAACCACATCCTCAACAACCTCGTGGAAGCATACCTCATCTCAGCATCAGACAAGAGT 1200

Qy   1201 CGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTG 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTG 1260

Qy   1261 CAGCCCAAAGTGACCGTCTTTTTCTGATGAAGAAGGAGTTCAGAGGACCTGCTGGAG 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CAGCCCAAAGTGACCGTCTTTTTCTGATGAAGAAGGAGTTCAGAGGACCTGCTGGAG 1320

Qy   1321 CTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTTGTGCCGGCAG 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTTGTGCCGGCAG 1380

Qy   1381 TGTCCTGAGTACAGAAGGCAGGCGGCGGCGCCTCCCACTGCCCAGCACCCGAGGGCGAG 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TGTCCTGAGTACAGAAGGCAGGCGGCGGCGCCTCCCACTGCCCAGCACCCGAGGGCGAG 1440

Qy   1441 CCAGGAGCCCCAGCCCTGGGGATGCACCCTCAGCCTGACGACAGCA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CCAGGAGCCCCAGCCCTGGGGATGCACCCTCAGCCTGACGACAGCA 1500

Qy   1501 GTCCAGGATTACGTGTGCCCTGCAAGGAAGCCACGCCCTGTGCACCTGCTTCCAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GTCCAGGATTACGTGTGCCCTGCAAGGAAGCCACGCCCTGTGCACCTGCTTCCAG 1560

Qy   1561 CCCATGCCCGACCGGAGCCGGAGCCGAGCGGCGACCCGGCGTGCACCCTCAGCAGTGT 1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CCCATGCCCGACCGGAGCCGGAGCCGAGCGGCGACCCGGCGTGCACCCTCAGCAGTGT 1620

Qy   1621 GCGGTCGCTGCCCGGGGCGAGGGGCGATCCCCGTCGTCGCCCGCCACTCAGGTG 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GCGGTCGCTGCCCGGGGCGAGGGGCGATCCCCGTCGTCGCCCGCCACTCAGGTG 1680

```
Qy   1681 GGTGCCTGGCCCCGTTTGTGAGCTCAACCTGGTGACAAGTGTCTGGACGGCGTGCTG 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GGTGCCTGGCCCCGTTTGTGAGCTCAACCTGGTGACAAGTGTCTGGACGGCGTGCTG 1740

Qy   1741 AACAACAACAGCTACGAGCTCAGACATCCTGAAGAATTACCTGCCAACCAGAGGTTTGACA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AACAACAACAGCTACGAGCTCAGACATCCTGAAGAATTACCTGCCAACCAGAGGTTTGACA 1800

Qy   1801 TGGAAAAACATGTTGACCGAGAGCCTCGTGGCCTCTCCACGGGGAGTGTTTCTGCTGTCT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGGAAAAACATGTTGACCGAGAGCCTCGTGGCCTCTCCACGGGGAGTGTTTCTGCTGTCT 1860

Qy   1861 GATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGGTGGCCTGGCGCAGCTTCCGT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGGTGGCCTGGCGCAGCTTCCGT 1920

Qy   1921 GAGCTGACCTATCAGTAGTCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GAGCTGACCTATCAGTAGTCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACA 1980

Qy   1981 TCCCGTCCTGACTGCTACTGGGGCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCC 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 TCCCGTCCTGACTGCTACTGGGGCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCC 2040

Qy   2041 ATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCT 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 ATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCT 2100

Qy   2101 GAGCAGCTTTCAGCACTGAGGTGAAGAGAGCGTGTTTTAAAATACAGAGACAAGCAGACG 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GAGCAGCTTTCAGCACTGAGGTGAAGAGAGCGTGTTTTAAAATACAGAGACAAGCAGACG 2160

Qy   2161 TCAAGGTGTTTTCACAGCCCCCTCGAGGGAGTGGACGCTCCGACACAGTGCTCTGG 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TCAAGGTGTTTTCACAGCCCCCTCGAGGGAGTGGACGCTCCGACACAGTGCTCTGG 2220

Qy   2221 GGTGACTCTTCTGTGGAGCTTTTTACCCTTCCGAGTGAGACCCTCCCCAGAGCCCGGGGG 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GGTGACTCTTCTGTGGAGCTTTTTACCCTTCCGAGTGAGACCCTCCCCAGAGCCCGGGGG 2280

Qy   2281 CCGCAGCCCGCCCCTCGTGGTGGGCTGGCAGGGCTCGTGGTGGTGGCATCAGCAGCAGAGA 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CCGCAGCCCGCCCCTCGTGGTGGGCTGGCAGGGCTCGTGGTGGTGGCATCAGCAGCAGAGA 2340

Qy   2341 CGAAGCCTTTCTGTAACATGCGGCCGTCCCGCCGAGAGGGGCAGTTTTGCTCTTTTGTAC 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CGAAGCCTTTCTGTAACATGCGGCCGTCCCGCCGAGAGGGGCAGTTTTGCTCTTTTGTAC 2400

Qy   2401 ATTTTCCGAAACTACAGTTAAAGCAGCAGAAGTCTGTTTTCAAGGGAGAAGTTCAAGGGAGAAG 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 ATTTTCCGAAACTACAGTTAAAGCAGCAGAAGTCTGTTTTCAAGGGAGAAGTTCAAGGGAGAAG 2460

Qy   2461 GGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTACAGGA 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTACAGGA 2520

Qy   2521 CGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTTTTATTTTAATA 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTTTTATTTTAATA 2580

Qy   2581 GGTTTGGTGCTTCATCTTCTAATAGATTTAAATGTCACAAACTGTAGCACAAATAATATA 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 GGTTTGGTGCTTCATCTTCTAATAGATTTAAATGTCACAAACTGTAGCACAAATAATATA 2640

Qy   2641 ATTTATAAATTTACAAATTGACAAAAAAAAAAAAAAA 2679
          |||||||||||||||||||||||||||||||||||||
Db   2641 ATTTATAAATTTACAAATTGACAAAAAAAAAAAAAAA 2679
```

RESULT 2
AAH14542
ID   AAH14542 standard; cDNA; 2639 BP.

XX
AC   AAH14542;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:12099.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0024B036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000;   2000JP-00118776.
PR   02-MAY-2000;   2000JP-00183767.
PR   09-JUN-2000;   2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 12099; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAH13629 to AAH13632 represent
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 2639 BP; 620 A; 706 C; 760 G; 553 T; 0 U; 0 Other;

Query Match   79.6%; Score 2133; DB 4; Length 2639;
Best Local Similarity   100.0%; Pred. No. 0;
Matches 2133; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    529 GATACCTCAGGTGCAGGTGCAGGGCGAGGGGCCGATCCCCGGTGCCTCCCGTCGTCGCGCC 588
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GATACCTCAGGTGCAGGTGCAGGGCGAGGGGCCGATCCCCGGTGCCTCCCGTCGTCGCGCC 540

Qy    589 GCCACTCAGGTGTGCTTTGAGGAACCACCAGCCATCAACATCGACGTCAGACCCTCTTCCCC 648
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCCACTCAGGTGTGCTTTGAGGAACCACCAGCCATCAACATCGACGTCAGACCCTCTTCCCC 600
```

```
Qy   649 ACAGCCTCGGCCTCTCCACGGAGCCTTCCTGCAGGGCCGAGACGCGTTCCTCCAGTTGT  708
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ACAGCCTCGGCCTCTCCACGGAGCCTTCCTGCAGGGCCGAGACGCGTTCCTCCAGTTGT  660

Qy   709 GGGTCTCGGGGTGGTGGCATCTCCCCTAAAGGAAGTGGTCCCTCGTGGCAAGTGATGAA  768
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GGGTCTCGGGGTGGTGGCATCTCCCCTAAAGGAAGTGGTCCCTCGTGGCAAGTGATGAA  720

Qy   769 GTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAAGACTCGCGTCGTTTCGTCGTTGGAA  828
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAAGACTCGCGTCGTTTCGTCGTTGGAA  780

Qy   829 CCCCAGGATCAGGAGGATTTGGAGCCCGTGAACAACCGGTAGAAATGCCCAAACCGTCCACGGAG  888
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 CCCCAGGATCAGGAGGATTTGGAGCCCGTGAACAACCGGTAGAAATGCCCAAACCGTCCACGGAG  840

Qy   889 GACCTGAACGGGCAGTTGTTGGTCGCACGACTCGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCG  948
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 GACCTGAACGGGCAGTTGTTGGTCGCACGACTCGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCG  900

Qy   949 GACCGCAGACAGCGCGGCTGGAAGCTGGAAGCTGGAAGATGAGGACGCGTCGACATGCATCATC 1008
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GACCTGAGACAGCGCGGCTGGAAGCTGGAAGCTGGAAGATGAGGACGCGTCGACATGCATCATC  960

Qy  1009 TGCCAGGACCTGCTGCACGACTCGTGAGTTTGCAGCCCTGTCCGTCCCTGTGTCCGCG 1068
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TGCCAGGACCTGCTGCACGACTCGTGAGTTTGCAGCCCTGTCCGTCCCTGTGTCCGCG 1020

Qy  1069 GCTTGCTACTCGGGCTGACGATGGAGCGCGTCGTCCCTGTGTCCTGTCCTGGAAGCATACCTCATCCAGCAT 1128
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 GCTTGCTACTCGGGCTGACGATGGAGCGCGTCGTCCCTGTGTCCTGTCCTGGAAGCATACCTCATCCAGCAT 1080

Qy  1129 GAGCGGATCTGTAAAAACCACATCCTCAACAACCTCGTCGGAAGCATACCTCCTCATCCATCCAGCAT 1188
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GAGCGGATCTGTAAAAACCACATCCTCAACAACCTCGTCGGAAGCATACCTCCTCATCCATCCAGCAT 1140

Qy  1189 CCAGAACAAGAGTCGCAGTGAAGAAGATGTGCAAAGATGGATGCCAGGAATAAAATCACT 1248
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CCAGAACAAGAGTCGCAGTGAAGAAGATGTGCAAAGATGGATGCCAGGAATAAAATCACT 1200

Qy  1249 CAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTCTGATGAAGAAGGAGTTCAGAG 1308
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTCTGATGAAGAAGGAGTTCAGAG 1260

Qy  1309 GACCTGCTGGAGCGTGTCAGACGTTGACACGTGAGTCCTCAGACATTAGCCAGCCATACGTC 1368
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 GACCTGCTGGAGCGTGTCAGACGTTGACACGTGAGTCCTCAGACATTAGCCAGCCATACGTC 1320

Qy  1369 GTGTGCCGCGCAGTGTCCAGCCATGCCGGCGCAGCCGGCAGCCTCCCCAGCA 1428
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 GTGTGCCGCGCAGTGTCCAGCCATGCCGGCGCAGCCGGCAGCCTCCCCAGCA 1380

Qy  1429 CCCGAGGGCGGAGCCCACAGGCCCTGGCCCCGTTTGTGAGCTCAACCTCAGTGTCTG 1488
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 CCCGAGGGCGGAGCCCACAGGCCCTGGCCCCGTTTGTGAGCTCAACCTCAGTGTCTG 1440

Qy  1489 CTGACGACAGCAGTCCAGGATTACGTGTGCCCCTCGCAAGGAAGCCACGCCCTGTGCACC 1548
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 CTGACGACAGCAGTCCAGGATTACGTGTGCCCCTCGCAAGGAAGCCACGCCCTGTGCACC 1500

Qy  1549 TGCTGCTTCCAGCCATGCCGGACCCGAGCGGCCGAGAGCCGGAGGCGTGTCGCC 1608
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 TGCTGCTTCCAGCCATGCCGGACCCGAGCGGCCGAGAGCCGGAGGCGTGTCGCC 1560

Qy  1609 CCTTCAGCAGTGTGCGGTCTGCCTCTGCAGCCTTTCTGCCACCTGTACTGGGGGCTCGCACCCG 1668
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 CCTTCAGCAGTGTGCGGTCTGCCTCTGCAGCCTTTCTGCCACCTGTACTGGGGGCTGCACCCG 1620

Qy  1669 ACCGGCTGCTACGGGCTGCCTGGCCCCCGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTG 1728
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 ACCGGCTGCTACGGGCTGCCTGGCCCCCGTTTGTGAGCTCAACCTGGGTGCACAAGTGTCTG 1680
```

```
Qy  1729 GACGGCGTGCTGAACAACAACAGCTACGACGTCAGACATCCTGAAGAATTACCTGGCAACC 1788
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 GACGGCGTGCTGAACAACAACAGCTACGACGTCAGACATCCTGAAGAATTACCTGGCAACC 1740

Qy  1789 AGAGTTTGACATGAAAAACATGTGACCGGAGAGCCTCGTGGCTCTCCAGCGGGGGAGTG 1848
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 AGAGTTTGACATGAAAAACATGTGACCGGAGAGCCTCGTGGCTCTCCAGCGGGGGAGTG 1800

Qy  1849 TTTCTGCTGTCTGATTACAGAGTCACGGAGACACCGTCTGTGTACTGCTGTCGTGGCCTG 1908
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 TTTCTGCTGTCTGATTACAGAGTCACGGAGACACCGTCTGTGTACTGCTGTCGTGGCCTG 1860

Qy  1909 CGCAGCTTCCGTGAGCTGACCTATCAGTAGTCGGCAGAACATTCCTGCTTCCGAGTTGCCA 1968
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 CGCAGCTTCCGTGAGCTGACCTATCAGTAGTCGGCAGAACATTCCTGCTTCCGAGTTGCCA 1920

Qy  1969 GTGGCCGTAACATCCCGTCCTGACTGCTACTGGGGCGTAACTGCCGCACTCAGGTGAAA 2028
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 GTGGCCGTAACATCCCGTCCTGACTGCTACTGGGGCGTAACTGCCGCACTCAGGTGAAA 1980

Qy  2029 GCTCACCACCACCGCCATGAAATTCAATCATCATATCTGTGAACACGACAAGGTTCAAAAACTAAGCA 2088
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 GCTCACCACCACCGCCATGAAATTCAATCATCATATCTGTGAACACGACAAGGTTCAAAAACTAAGCA 2040

Qy  2089 TCCAGAGGCCCTGAGCAGTTTCAGCAGTTTGTTTCACAGCCCCCTGAGGGAAGGACGCAGGGTCTCCGA 2148
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 TCCAGAGGCCCTGAGCAGTTTCAGCAGTTTGTTTCACAGCCCCCTGAGGGAAGGACGCAGGGTCTCCGA 2100

Qy  2149 GAGACAAGCACGTCAAGGTGTTTCACAGCCCCTGTGGAGTCCTCTGTGAGCGGCTGGGCAGCCCTCCCCA 2208
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 GAGACAAGCACGTCAAGGTGTTTCACAGCCCCTGTGGAGTCCTCTGTGAGCGGCTGGGCAGCCCTCCCCA 2160

Qy  2209 CAGGTGCTCTGGGGTGACTCTTCGTCGTGGAGCTGGGCAGGGCTCGTGGTGGCAT 2268
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 CAGGTGCTCTGGGGTGACTCTTCGTCGTGGAGCTGGGCAGGGCTCGTGGTGGCAT 2220

Qy  2269 GAGCCCCGGGGGCGGCCGCAGCCCGCCCCGTTCTGTAACATGCCGCCGTCCGAGAGGGGCAGTTTT 2328
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 GAGCCCCGGGGGCGGCCGCAGCCCGCCCCGTTCTGTAACATGCCGCCGTCCGAGAGGGGCAGTTTT 2280

Qy  2329 CAGCAGCAGAGACGAAGCCTTTCGTAACATCAGTTAAAGCAGAAGCTGTTTCAGGAAAAAGTT 2388
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 CAGCAGCAGAGACGAAGCCTTTCGTAACATCAGTTAAAGCAGAAGCTGTTTCAGGAAAAAGTT 2340

Qy  2389 GCTCTTTGTACATTTCCGAAACTACAGTTTATCAAAAACATTGTTCGTGCGGGAGCCATAAGTTTA 2448
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 GCTCTTTGTACATTTCCGAAACTACAGTTTATCAAAAACATTGTTCGTGCGGGAGCCATAAGTTTA 2400

Qy  2449 TCAAGGGAGAAGGGCCAAGTTTATCATGTGCTTATCTTCTAAATAAGATTTAAATGTCACAAACTGTAGC 2508
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TCAAGGGAGAAGGGCCAAGTTTATCATGTGCTTATCTTCTAAATAAGATTTAAATGTCACAAACTGTAGC 2460

Qy  2509 CAGCCTCACAGACGTACACAATATCCTGCTGTCGGGAAAAACCACAGCATTTATCTCATTT 2568
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 CAGCCTCACAGACGTACACAATATCCTGCTGTCGGGAAAAACCACAGCATTTATCTCATTT 2520

Qy  2569 TTTATTTTAATAGGTTTGGTGCTTATCTTCTAAATAAGATTTAAATGTCACAAACTGTAGC 2628
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 TTTATTTTAATAGGTTTGGTGCTTATCTTCTAAATAAGATTTAAATGTCACAAACTGTAGC 2580

Qy  2629 ACAAATAATATAATTAATTATAATTTACAAATTGAC 2661
         ||||||||||||||||||||||||||||||||||||
Db  2581 ACAAATAATATAATTAATTATAATTTACAAATTGAC 2613
```

```
DE   Nucleotide sequence of a human ring finger protein designated FHAR1.
XX
KW   FHAR1; RING finger protein; cancer; vaccine; ss.
XX
OS   Homo sapiens.
XX
FH            Key            Location/Qualifiers
FT   CDS            1..1995
FT                  /tag= a
FT                  /product= "ring finger protein FHAR1"
XX
PN   WO200142430-A1.
XX
PD   14-JUN-2001.
XX
PF   07-DEC-2000; 2000WO-US033094.
XX
PR   08-DEC-1999; 99US-00456876.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Zhou B,  Zhu Y,  Chaturvedi P,  Hurle MR,  Li X;
XX
DR   WPI; 2001-381663/40.
DR   P-PSDB; AAB83843.
XX
PT   New FHAR1 polypeptide, a member of the RING finger protein family for
PT   diagnosing and treating cancer, and for use in anti-cancer vaccines.
XX
PS   Claim 2; Page 18-19; 28pp; English.
XX
CC   The present sequence encodes a FHAR1 polypeptide, which is a member of
CC   the RING finger protein family. FHAR1 is useful in the treatment of
CC   cancer, and as a vaccine for inducing an immunological response in a
CC   mammal. FHAR1 polynucleotides may also be used as a diagnostic reagent
CC   through detection of mutations in the associated gene, and for chromosome
CC   localization studies, and tissue expression studies. FHAR1 antibodies are
CC   useful to isolate and to identify clones expressing the polypeptides, or
CC   to purify the polypeptides by affinity chromatography and to treat cancer
XX
SQ   Sequence 1995 BP; 465 A; 563 C; 580 G; 387 T; 0 U; 0 Other;

Query Match       74.5%;  Score 1995;  DB 5;  Length 1995;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 1995;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     91 ATGGAGCGGCGCCGAGGAAGGCAAGCAGTGCCGCGCCGCCGCCGGGACGGCGGCTCCTG 150
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGAGCGGCGCCGAGGAAGGCAAGCAGTGCCGCGCCGCCGCCGGGACGGCGGCTCCTG  60

Qy    151 CGTCTGGGCGCGGAGGAGGGGGCAGCCGCACGTCTCCTGAGGAAGCGGGAGTGGACCATC 210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGTCTGGGCGCGGAGGAGGGGGCAGCCGCACGTCTCCTGAGGAAGCGGGAGTGGACCATC 120

Qy    211 GGGCGGCGAGACCAGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGATCAC 270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGGCGGCGAGACCAGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGATCAC 180

Qy    271 TGTAGAATTGTAGTGGATGAAAAATCAGGTCAGGTTGTTAAGAAGCACTGGAGATACCAGCACCAGT 330
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGTAGAATTGTAGTGGATGAAAAATCAGGTCAGGTTGTTAAGAAGCACTGGAGATACCAGCACCAGT 240

Qy    331 GGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGACATGCCCTTTACAGACTGGG 390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGACATGCCCTTTACAGACTGGG 300

Qy    391 GATGTCATCTCACTTGGTGTACAGGAAGAATGAACCGGAACACAACGTGGCATACCTCTAT 450
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GATGTCATCTCACTTGGTGTACAGGAAGAATGAACCGGAACACAACGTGGCATACCTCTAT 360

Qy    451 GAATCTTTAAGTGAAAAGCAAGCAAGCATGACACAAGAATCCTTTGAAGCTAACAAGGAAAAT 510
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GAATCTTTAAGTGAAAAGCAAGCAAGCATGACACAAGAATCCTTTGAAGCTAACAAGGAAAAT 420

Qy    511 GTGTTCCATGGGACCAAAGATACCTCAGGTGCAGGTGCCGATCCCCGG 570
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTGTTCCATGGGACCAAAGATACCTCAGGTGCAGGTGCCGATCCCCGG 480

Qy    571 GTCCCTCCGTCGTCGCCGCCGCCACTCAGGTGTGCTTTGAGGAACCACAGCCATCAACATCG 630
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GTCCCTCCGTCGTCGCCGCCGCCACTCAGGTGTGCTTTGAGGAACCACAGCCATCAACATCG 540

Qy    631 ACGTCAGACCTCTTCCCGGCCTCGGCCTCGGCCTCGGCCTTCTCCTGCAGGGGCGA 690
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACGTCAGACCTCTTCCCGGCCTCGGCCTCGGCCTCGGCCTTCTCCTGCAGGGGCGA 600

Qy    691 GAGCGTTCCTCCAGTTGTGGGTCTGGGTGGCATCTCCCTAAAGGAAGTGGTCCC 750
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAGCGTTCCTCCAGTTGTGGGTCTGGGTGGCATCTCCCTAAAGGAAGTGGTCCC 660

Qy    751 TCTGTGGCAAGTGATGAAGTCTCCAGCTTGCCTCAGCTCTCCAGACAGAAAGACTGCG 810
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TCTGTGGCAAGTGATGAAGTCTCCAGCTTGCCTCAGCTCTCCAGACAGAAAGACTGCG 720

Qy    811 TCCTTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAAATG 870
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TCCTTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAAATG 780

Qy    871 AGAGGAGATGGGGACCTTGACCTGAACGGCACCAGTTGTTGGTTCGGAAGCCAGCAAGATGGAGGAG 930
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AGAGGAGATGGGGACCTTGACCTGAACGGCACCAGTTGTTGGTTCGGAAGCCAGCAAGATGGAGGAG 840

Qy    931 GCCCAAACCGTCCACGAGGACGTCAGAGCAGCGGCTGGGAAGCCAGCCAGACAAGATGGAGGAG 990
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GCCCAAACCGTCCACGAGGACGTCAGAGCAGCGGCTGGGAAGCCAGCCAGACAAGATGGAGGAG 900

Qy    991 ACGTGACATGCATCATCTGCCAGGACTCTGCTCGGTAGTTGCAGCCCTGC 1050
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ACGTGACATGCATCATCTGCCAGGACTCTGCTCGGTAGTTGCAGCCCTGC 960

Qy   1051 ATGCACACGTTCTGCGCGGCTTGCTACTCGGGCTGGATGGAGGCTCGTCCCTGTCCT 1110
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 ATGCACACGTTCTGCGCGGCTTGCTACTCGGGCTGGATGGAGGCTCGTCCCTGTCCT 1020

Qy   1111 ACCTGCGCGTGTCCCGTGGAGCGGATCGTGAAAAACCACATCCTCAACAACCTCGTGGAA 1170
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ACCTGCGCGTGTCCCGTGGAGCGGATCGTGAAAAACCACATCCTCAACAACCTCGTGGAA 1080

Qy   1171 GCATACCTCATCCAGACAAGAGTCGAGTGAAGAAGATGTGCAAAGTATGGAT 1230
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GCATACCTCATCCAGACAAGAGTCGAGTGAAGAAGATGTGCAAAGTATGGAT 1140

Qy   1231 GCCAGGAATAAAATCACTCAAGACATGCCTGAGGACGCGGAAGCGGGAGTGGACCATC 1290
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GCCAGGAATAAAATCACTCAAGACATGCCTGAGGACGCGGAAGCGGGAGTGGACCATC 1200

Qy   1291 GAAGAAGGGAGTTCAGAGGACCTGCTGGAGACGTTGACAGTGTCAGTGAGTCTTCAGAC 1350
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GAAGAAGGGAGTTCAGAGGACCTGCTGGAGACGTTGACAGTGTCAGTGAGTCTTCAGAC 1260

Qy   1351 ATTAGCCAGCCATACGTCGTGTGCCGAGGGGCGAGCCAAGGAGCCCTGGGGGATGCA 1410
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ATTAGCCAGCCATACGTCGTGTGCCGAGGGGCGAGCCAAGGAGCCCTGGGGGATGCA 1320

Qy   1411 CCTCCCACTGCCCAGCACCCGAGGGCGGAGCCAGGAGCCCACAGACTGGG 1470
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CCTCCCACTGCCCAGCACCCGAGGGCGGAGCCAGGAGCCCACAGACTGGG 1380

Qy   1471 CCCTCCACGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTTGTGTCCCTCTGCAAGGA 1530
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CCCTCCACGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTTGTGTCCCTCTGCAAGGA 1440

Qy   1531 AGCCACGCCCGTGCACCTGCTTCCAGCCCATGCCCGACCCGGACCGGCGAGCGGGAGCGGAG 1590
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AGCCACGCCCGTGCACCTGCTTCCAGCCCATGCCCGACCCGGACCGGCGAGCGGGAGCGGAG 1500
```

PS   Claim 1; SEQ ID NO 57; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC   and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC   rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a coding sequence of the
CC   invention
XX
SQ   Sequence 3181 BP; 756 A; 830 C; 879 G; 716 T; 0 U; 0 Other;

Query Match   60.4%; Score 1618; DB 6; Length 3181;
Best Local Similarity   99.7%; Pred. No.0;
Matches 1948; Conservative   0; Mismatches   0; Indels   3; Gaps   1;

Qy   709 GGGTCTGGGGGTGGTGGCATCCCCCTAAAGGAAGTGGTCCCTCTGCCAAGTGATGAA 768
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   591 GGGTCTGGGGGTGGTGGCATCCCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAA 650

Qy   769 GTCTCCAGCTTGCCTCAGCTCCCAGACAGAAAGACTGCGTCCTTTCGTCGTTGGAA 828
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   651 GTCTCCAGCTTGCCTCAGCTCCCAGACAGAAAGACTGCGTCCTTTCGTCGTTGGAA 710

Qy   829 CCCCAGGATCAGGAGGATTTGGAGCCGTGAAGAAGAGGAGGATGGGGACCTT 888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   711 CCCCAGGATCAGGAGGATTTGGAGCCGTGAAGAAGAGGAGGATGGGGACCTT 770

Qy   889 GACCTGAACGGGCAGTTGTTGTGGTCGGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAG 948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   771 GACCTGAACGGGCAGTTGTTGTGGTCGGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAG 830

Qy   949 GACGTCAGAGCAGCGCGTGGAAGCCAGACAGCAGCAAGATGGAGGAGACCGTGACATGCATCATC 1008
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   831 GACGTCAGAGCAGCGCGTGGAAGCCAGACAGCAGCAAGATGGAGGAGACCGTGACATGCATCATC 890

Qy   1009 TGCCAGACCTGCTGCGACGACTGCGTGAGTTGCAGCCCTGCATGCACACGTTCTGCGCG 1068
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   891 TGCCAGACCTGCTGCGACGACTGCGTGAGTTGCAGCCCTGCATGCACACGTTCTGCGCG 950

Qy   1069 GCTTGTACTGGGCTGGATGGAGCGCTCGTCCCGTCCTACCTGCGCCGTGCCCGTG 1128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   951 GCTTGTACTGGGCTGGATGGAGCGCTCGTCCCGTCCTACCTGCGCCGTGCCCGTG 1010

Qy   1129 GAGCGGATCTGTAAAAACCATCCTCACAACCTCGTGGAAGCATACCTCATCCAGCAT 1188
          ||||||||||||||||| |||| ||||||||||||||||||||||||||||||||||
Db   1011 GAGCGGATCTGTAAAAACCACATCCTCAACAACCTCGTGGAAGCATACCTCATCCAGCAT 1070

Qy   1189 CCAGACAAGAGTCGCAGTGAAGAAGAGTGCAAGTATCGGATGCCAAGAATAAAATCACT 1248
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1071 CCAGACAAGAGTCGCAGTGAAGAAGAGTGCAAGTATCGGATGCCAAGAATAAAATCACT 1130

Qy   1249 CAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTCTGATGAAGAAGGAGTTCAGAG 1308
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1131 CAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTCTGATGAAGAAGGAGTTCAGAG 1190

Qy   1309 GACCTGCTGGAGCTGTGACAGTTGACAGTGAGTCCTCAGACATTAGCAGCCATACGTC 1368
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1191 GACCTGCTGGAGCTGTGACAGTTGACAGTGAGTCCTCAGACATTAGCAGCCATACGTC 1250

Qy   1369 GTGTGCCGGCAGTGTCCTGAGTACAGAAGCAGGCGGCGGCAGCCTCCCACTGCCCAGCA 1428
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1251 GTGTGCCGGCAGTGTCCTGAGTACAGAAGCAGGCGGCGGCAGCCTCCCACTGCCCAGCA 1310

Qy   1429 CCCGAGGGGCGGAGCCAGGAGCCCCAGGCCCCAGGCCCTGGGGGATGCACCCTCCCAGTCGGTCAGC 1488
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1311 CCCGAGGGGCGGAGCCAGGAGCCCCAGGCCCCAGGCCCTGGGGGATGCACCCCACGTCGGTCAGC 1370

Qy   1489 CTGACCGACAGCAGTCAGGATTACGTGTGCCCTCTGCAAGGAAGCACGCACCGCCCTGCACC 1548

Qy   1591 CAGGACCCGCGTGCGCCCTCCAGCAGTGTGGGTCGCCTGCAGCCTTTCTGCCACCTG 1650
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CAGGACCCGCGTGCGCCCCTCCAGCAGTGTGCGGTCGCCTGCCTTTCTGCCACCTG 1560

Qy   1651 TACTGGGGTGACACCCGGACCGGCTGCTACGGCTGCTGCCTGGCCCGTTTTGTGAGCTCAAC 1710
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TACTGGGGGTGCACCCGGACCGGCTGCTACGGCTGCTGCCTGGCCCCGTTTTGTGAGCTCAAC 1620

Qy   1711 CTGGGTGACAAGTGTCTGGACGCGTGCTGACAACAACAGCAGCTACGAGTCAGAGCATCCTG 1770
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CTGGGTGACAAGTGTCTGGACGCGTGCTGACAACAACAGCTACGAGTCAGAGCATCCTG 1680

Qy   1771 AAGAATTACCTGGCAACCAGAGAGTTTGACATGGAAAAACATGTTGACCGGAGAGCCTCGTG 1830
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 AAGAATTACCTGGCAACCAGAGAGTTTGACATGGAAAAACATGTTGACCGGAGAGCCTCGTG 1740

Qy   1831 GCTCTCCAGCGCGGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTG 1890
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GCTCTCCAGCGCGGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTG 1800

Qy   1891 TGTTACTGCTGTGCCTGCGCAGCTTCCGTGACCTATCAGTATCGGCAGAACATT 1950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGTTACTGCTGTGCCTGCGCAGCTTCCGTGACCTATCAGTATCGGCAGAACATT 1860

Qy   1951 CCTGCTTCCGAGTTGCCAGTGGCCGGTAACATCCCGTCCTGACTGTACTGGGGGCCGTAAC 2010
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CCTGCTTCCGAGTTGCCAGTGGCCGGTAACATCCCGTCCTGACTGTACTGGGGGCCGTAAC 1920

Qy   2011 TGCCGCACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATATCTGTGACACAGACA 2070
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TGCCGCACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATATCTGTGACACAGACA 1980

Qy   2071 AGGTTCAAAAACTAA 2085
          |||||||||||||||
Db   1981 AGGTTCAAAAACTAA 1995

RESULT 4
ABN9646
ID   ABN9646 standard; cDNA; 3181 BP.
XX
AC   ABN59646;
XX
DT   28-JUN-2002 (first entry)
XX
DE   Novel human coding sequence SEQ ID NO: 57.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US026015.
XX
PR   11-SEP-2000; 2000US-00659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Weihman T, Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   P-PSDB; ABB97233.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis.
XX

Db   2449 TTTATTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGC 2508
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   2629 ACAAATAATATATA|||||||||||||||||||||||||||||||||||||||| 2661
Db   2509 ACAAATAATATAATTATTATAATTTACAAATTGAC 2541

RESULT 5
AAH14556
ID   AAH14556 standard; cDNA; 3138 BP.
XX
AC   AAH14556;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:12127.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   BP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999; 99JP-00248036.
PR   27-AUG-1999; 99JP-00300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 12127; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 3138 BP; 738 A; 824 C; 873 G; 703 T; 0 U; 0 Other;

Query Match   54.7%; Score 1465; DB 4; Length 3138;

Qy   1371 CTGACGACAGCAGTCCAGGATTACGTGCTGCCCTCTGCAGAAGAAGCCACGCCCTGTGCACC 1430
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   1549 TGCTGCTTCCAGCCATGCCCGACCGGAGAGCCGAGCCGAGCCCGCGTGTCGCC 1608
           ||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   1431 TGCTGCTTCCAGCCCATGCCCGACCGGAGAGCCGAGCCGAGCCCGCGTGTCGCC 1490

Qy   1609 CCTCAGCAGTGTGCGGTCTGCTGCCACCTGTACTGGCGGTCGCACCCGG 1668
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1491 CCTCAGCAGTGTGCGGTCTGCTGCCACCTGTACTGGCGGTCGCACCCGG 1550

Qy   1669 ACCGGCGTGCTACGGCTGCTGCCGCCCCGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTG 1728
           |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1551 ACCGGCGTACGGCTGCTGCCGCCCCGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTG 1610

Qy   1729 GACGGCGGTGCTGAACACAACAGCTACGAGTCAGACATCCTGAAGAATTACCTGGCAACC 1788
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1611 GACGGCGGTGCTGAACACAACAGCTACGAGTCAGACATCCTGAAGAATTACCTGGCAACC 1670

Qy   1789 AGAGGTTTGACATGGAAAAACATGTTGACCGAGAGCCTCTTCGTGCTCTTCCAGCGGGGAGTG 1848
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1671 AGAGGTTTGACATGGAAAAACATGTTGACCGAGAGCCTCTTCGTGCTCTTCCAGCGGGGGAGTG 1730

Qy   1849 TTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGGTTCTGTTACTGCTGTGGGCCTG 1908
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1731 TTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGGTTCTGTTACTGCTGTGGGCCTG 1790

Qy   1909 CGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCA 1968
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   1791 CGCAGCTTCCGTGAGCTGACCTATCAGTATCAGCAGAACATTCCTGCTTCCGAGTTGCCA 1850

Qy   1969 GTGGCCGTAACATCCCGTCCTGACTGGGCCCGTAACTGCCCGCCACTCAGGTGAAA 2028
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1851 GTGGCCGTAACATCCCGTCCTGACTGGGCCCGTAACTGCCCGCCACTCAGGTGAAA 1910

Qy   2029 GCTCCACCACGCCATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCA 2088
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1911 GCTCCACCACGCCATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCA 1970

Qy   2089 TCCAGAGGCCCTGAGCACTGGAGGTGAGGAGAGAGAGAAAATACA 2148
           |||||||||||||||||||||||||||| |||||||||||| ||
Db   1971 TCCAGAGGCCCTGAGCACTGGAGGTGTGTTTTTAAAATACC 2030

Qy   2149 GAGACAAGCACGTCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGGACGCAGGGTCTCCGA 2208
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2031 GAGACAAGCACGTCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGGACGCAGGGTCTCCGA 2090

Qy   2209 CAGGTGCTCTGGGGTGACTCCTTCTGTGGAGCTTTTACCCTCGAGTGAGACCCTCCCCA 2268
           |||||||||||||||||||||||||||||||| |||||||| ||||||||||||||||
Db   2091 CAGGTGCTCTGGGGTGACTCCTTCTGTGGAGC--TTTACCCTCTGAGTGAGACCCTCCCCA 2148

Qy   2269 GAGCCCCGGGGGGCCCAGCCCGCCGCCCCGCCCGTCCGGTGGTCGTGGTGGCAT 2328
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2149 GAGCCCCGGGGGGCCCAGCCCGCCGCCCCGCCCGTCCGGTGGTCGTGGTGGCAT 2208

Qy   2329 CAGCAGCAGAGCAGAAGCCTTTCGTGAAACATGCCGGCCGTCCCGCCGAGAGGGGCAGTTTT 2388
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2209 CAGCAGCAGAGAGCAGAAGCCTTTCGTGAAACATGCCGGCCGTCCCGCCGAGAGGGGCAGTTTT 2268

Qy   2389 GCTCTTTGTACATTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAGTT 2448
           ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   2269 GCTCTTTGTACATTTCCGAAACTACAGTTAAAGCGGAAGTCTGTTTTCAGGAAAAGTT 2328

Qy   2449 TCAAGGGAGGAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAGGAGGCATAAGTTTA 2508
           ||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   2329 TCAAGGAGGAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGAGGCATAAGTTTA 2388

Qy   2509 CAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACCAGCATTTATCTATTT 2568
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2389 CAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACCAGCATTTATCTATTT 2448

Qy   2569 TTTATTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGC 2628
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Best Local Similarity 99.6%; Pred. No. 0; Gaps 1;
Matches 1945; Conservative 0; Mismatches 6; Indels 2;

```
Qy    709 GGGTCTGGGGGTGGTGGCATCCCCTAAAGGAGTTGTCCCTCTGTGGCAAGTGATGAA 768
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    560 GGGTCTGGGGGTGGTGGCATCTCCCCTAAAGGAGTTGTCCCTCTGTGGCAAGTGATGAA 619

Qy    769 GTCTCCAGCTTGCCTCAGCTCTCCCGACAGAGAAGACTGCGTCCTTTTCGTCGTTGGAA 828
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    620 GTCTCCAGCTTGCCTCAGCTCTCCCGACAGAGAAGACTGCGTCCTTTTCGTCGTTGGAA 679

Qy    829 CCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAGAAAATGAGGAGGATGGGGACCTT 888
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    680 CCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAGAAAATGAGGAGGATGGGGACCTT 739

Qy    889 GACCTGAACGGGGCAGTTGTTGGTCGCACAACCGTAGAAATGCCCAAACCGTCCACGAG 948
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    740 GACCTGAACGGGGCAGTTGTTGGTCGCACAACCGTAGAAATGCCCAAACCGTCCACGAG 799

Qy    949 GACGTCAGAGCAGCGGCTGGGAAGCCAGACAAGAGGAGACGCTGACATGCATCATC 1008
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    800 GACGTCAGAGCAGCGCGGCTGGGAAGCCAGACAAGAGGAGACGCTGACATGCATCATC 859

Qy   1009 TGCCAGGACCTGCTGCACGACTGCGTGAGTTTGCAGCCCTGCATGCACACGGTTCTGCGCG 1068
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    860 TGCCAGGACCTGCTGCACGACTGCGTGAGTTTGCAGCCCTGCATGCACACGGTTCTGCGCG 919

Qy   1069 GCTTGCTACTCGGGCTGGATGAGCGCTCGTCCCGTGTCCTGTCCCACCTGCCGCGTTCCCGTG 1128
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    920 GCTTGCTACTCGGGCTGGATGAGCGCTCGTCCCGTGTCCTGTCCCACCTGCCGCGTTCCCGTG 979

Qy   1129 GAGCGGGATCGTGTAAAAACCACATTCCTCAACAACCTGCTGGAGCCATACCTCATCCAGCAT 1188
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    980 GAGCGGGATCGTGTAAAAACCACATTCCTCAACAACCTGCTGGAGCCATACCTCATCCAGCAT 1039

Qy   1189 CCAGACAAGAGTGCAGTGAAGAAGATGTCAAAGTATGGATCCAGGAATAAAATCACT 1248
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1040 CCAGACAAGAGTGCAGTGAAGAAGATGTCAAAGTATGGATCCAGGAATAAAATCACT 1099

Qy   1249 CAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTTCTGATGAAGAGGAGTTCAGAG 1308
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1100 CAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTTCTGATGAAGAGGAGTTCAGAG 1159

Qy   1309 GACCTGCTGGAGCGTGTCAGACGTTGACAGTGAGTCTCCTGACACATTAGCCGCCATAGTC 1368
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1160 GACCTGCTGGAGCGTGTCAGACGTTGACAGTGAGTCTCCTGACACATTAGCCGCCATAGTC 1219

Qy   1369 GTGTGCCGGCAGTCTCCTGAGTACAGAGGCAGGCGGCGCGCAGCCTCCCACTGCCCCAGCA 1428
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1220 GTGTGCCGGCAGTCTCCTGAGTACAGAGGCAGGCGGCGCGCAGCCTCCCACTGCCCCAGCA 1279

Qy   1429 CCCGAGGGGCGAGCCAGGAGCCCACAGCGCCCTGGGGGATGTCACCCTCCGTCCGTCAGC 1488
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1280 CCCGAGGGGCGAGCCAGGAGCCCACAGCGCCCTGGGGGATGTCACCCTCCGTCCGTCAGC 1339

Qy   1489 CTGACGACAGCAGTCCAGGATTACGTGTCCCCTGCGAAGGAGCCACGCCCCTGTGCACC 1548
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1340 CTGACGACAGCAGTCCAGGATTACGTGTCCCCTGCGAAGGAGCCACGCCCCTGTGCACC 1399

Qy   1549 TGCTGCTTCTTCCAGCCATGCCCGACCCGGAGCCGGAGCGCGAGCCGAGCACCGTGTCGCC 1608
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1400 TGCTGCTTCTTCCAGCCCATGCCCGACCCGGAGCCGGAGCGCGAGCCGAGCACCGTGTCGCC 1459

Qy   1609 CCTCAGCAGTGCGGGTGCGGTCTGCCTGGACTGGGGCTGCACCCGG 1668
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   1460 CCTCAGCAGTGCGGGTGCGGTCTGCCTGGACTGTACTGGGGCTGCACCCGG 1519

Qy   1669 ACCGGCTGCTACGGCGTACGGGCCCTGCCCTGGCCCCGTTTTGTGAGCTCAACCTGGGTGACAAGTGTCTG 1728
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1520 ACCGGCTGCTACGGCGTACGGGCCCTGCCCTGGCCCCGTTTTGTGAGCTCAACCTGGGTGACAAGTGTCTG 1579

Qy   1729 GACGGCGTGCTGAACAACAACAGCTACGACAGTCAGAGTCAGAGAATTACCTGGCAACC 1788
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1580 GACGGCGTGCTGAACAACAACAGCTACGACAGTCAGAGTCAGAGAATTACCTGGCAACC 1639

Qy   1789 AGAGGTTTGACATGGAAAAACATGTTGACCGAGCGTTCTCCAGCGGCGGGGAGTG 1848
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1640 AGAGGTTTGACATGGAAAAACATGTTGACCGAGCGTTCTCCAGCGGCGGGGAGTG 1699

Qy   1849 TTTCTGCTGCTGATTACAGAGTCACGGAGACACCGGTTCTGTGTTCTGTGTTCTGTGGCCTG 1908
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1700 TTTCTGCTGCTGATTACAGAGTCACGGAGACACCGGTTCTGTGTTCTGTGTTCTGTGGCCTG 1759

Qy   1909 CGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCA 1968
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1760 CGCAGCTTCCGTGAGCTGACCTATCAGTATCAGCAGAACATTCCTGCTTCCGAGTTGCCA 1819

Qy   1969 GTGGCCGTAACATCCCGTCCTGACTGCTACTGGGCGCTACTGCGCACTCAGGTGAAA 2028
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1820 GTGGCCGTAACATCCCGTCCTGACTGCTACTGGGCGCTACTGCGCACTCAGGTGAAA 1879

Qy   2029 GCTCACCACGCCATGAAAATCAATCATATCTGTGAACAGACAAGTTCAAAAACTAAGCA 2088
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1880 GCTCACCACGCCATGAAAATCAATCATATCTGTGAACAGACAAGTTCAAAAACTAAGCA 1939

Qy   2089 TCCAGAGGCCCTGAGCAGCTTCAGCACTGGAGGTGAAGAGAGCGTGTTTTTTAAAATACA 2148
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1940 TCCAGAGGCCCTGAGCAGCTTCAGCACTGGAGGTGAAGAGAGCGTGTTTTTTAAAATACA 1999

Qy   2149 GAGACAAGCACGTCAAGGTGTTTCACAGCCCCTGAGTGAGCAGGGCTCGTCGGTGGTGCAT 2208
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2000 GAGGCAAGCACGTCAAGGTGTTTCACAGCCCCTGAGTGAGCAGGGCTCGTCGGTGGTGCAT 2059

Qy   2209 CAGGTGCTCTGGGGTGACTCTTCTGTGGAGCTTTTCACCCTCTGAGTGAGACCCTCCCCA 2268
          ||||||||||||||||||||||||||||      |||||||||||||||||||||
Db   2060 CAGGTGCTCTGGGGTGACTCTTCTGTGGAGC--TTTACCCTCTGAGTGAGACCCTCCCCA 2117

Qy   2269 GAGCCCGGGGCCCGAGCCCGAGCCCGCCCTTCTGTAACATGCGGCCGTCCGAGAGGGCAGTTT 2328
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2118 GAGCCCGGGGCCCGAGCCCGAGCCCGCCCTTCTGTAACATGCGGCCGTCCGAGAGGGCAGTTT 2177

Qy   2329 CAGCAGCAGACGAGGAGCCCTTTCTGATGAAGAGGAGTTCAGAG 2388
          |||||||||||||||||||||||||||||||||||||||||||
Db   2178 CAGCAGCAGACGAGGAGCCCTTTCTGATGAAGAGGAGTTCAGAG 2237

Qy   2389 GCTCTTTTGTACATTTTCCGAAACTACACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAGTT 2448
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2238 GCTCTTTTGTACATTTTCCGAAACTACACAGTTAAAGCAGAAGTCTGTTTTTCAGGAAAAGTT 2297

Qy   2449 TCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAACGGGAGCATAAGTTTA 2508
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2298 TCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAACGGGAGCATAAGTTTA 2357

Qy   2509 CAGCCTACAGGACGTACACAATATCCTGCTCGTCGTGGGAAAAACCACAGCATTTTATCTATTT 2568
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2358 CAGCCTACAGGACGTACACAATATCCTGCTCGTCGTGGGAAAAACCACAGCATTTTATCTATTT 2417

Qy   2569 TTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGC 2628
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2418 TTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGC 2477

Qy   2629 ACAAATATATATAATTATAATTTACAAATTGAC 2661
          |||||||||||||||||||||||||||||||||
Db   2478 ACAAATATATATATTTTATAATTTACAAATTGAC 2510
```

RESULT 6
AAI88903/c
ID   AAI88903 standard; cDNA; 2259 BP.
XX
AC   AAI88903;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human polynucleotide SEQ ID NO 8963.
XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO2001164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
PR 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO08972.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
PS Claim 1; SEQ ID NO 8963; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell population. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2259 BP; 531 A; 612 C; 615 G; 501 T; 0 U; 0 Other;

Query Match   38.2%; Score 1024; DB 4; Length 2259;
Best Local Similarity   99.7%; Pred. No. 0;
Matches 1304; Conservative   0; Mismatches   2; Indels   2; Gaps   1;

Qy   1354 AGCGCAGCCATACGTCGTGTGTGCCCTCAGCAGTGTGCCGGCAGTACAGAAGGCAGGCGCGGCAGCCT 1413
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1965 AGCGCAGCCATACGTCGTGTGTGCCCTCAGCAGTGTGCCGGCAGTACAGAAGGCAGGCGCGGCAGCCT 1906

Qy   1414 CCCCACTGCCCAGCACCCGGAGGGCGAGCCAGGAGCGCCACAGCGCCCTGGGGGATGCACCC 1473
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1905 CCCCACTGCCCAGCACCCGGAGGGCGAGCCAGGAGCGCCACAGCGCCCTGGGGGATGCACCC 1846

Qy   1474 TCCACGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTGTGCCCTCTGCAAGGAAGC 1533
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1845 TCCACGTCCGTCAGCCTGCACCTGCTGCTCCAGCACCACAGCAGTCCAGGATTACGTGTGCCCTCTGCAAGGAAGC 1786

Qy   1534 CACGCCCTGTGCACCTGCTGCTGCTCTGCTTCCAGCCCATGCCCAGCCCCATGCCGAGAGCGGCGAGCAG 1593
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1785 CACGCCCTGTGCACCTGCTGCTGCTCTGCTTCCAGCCCATGCCCAGCCCCATGCCGAGAGCGGCGAGCAG 1726

Qy   1594 GACCCGCGTGCACCCGGACCCGGCTCGCCCCGTTGTGCGGCTCGGCGTCTGCGTCTGCTGCACCTG 1653
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1725 GACCCGCGTGCGCCCCGGACCCGGCTCGCCCCGTTTCTGCGAGCTCACCTGTAC 1666

Qy   1654 TGGGGCTGCACCCGGACCCGGACCCGGCTGCTACGGCGTCTGCTGCACCTG 1713
          |||||||||||||||||||||||||||||||||||||
Db   1665 TGGGGCTGCACCCGGACCCGGACCCGGCTGCTACGGCGTCTGCTGCACCTGTGAGCTCAACCTG 1606

Qy   1714 GGTGACAAGTGTCTGGACGGCGGCGTGCTGACACAACAGCTACGAGTCAGACATCCTGAAG 1773

Db   1605 GGTGACAAGTGTCTGGACGGCGGCGTGCTGACAACAACAGCTACGAGTCAGACATCCTGAAG 1546
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   1774 AATTACCTGGCAACCAGAGTTTGACATGGAGAAAACATGTTGACCGAGAGCCTCGTGCT 1833
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1545 AATTACCTGGCAACCAGAGTTTGACATGGAGAAAACATGTTGACCGAGAGCCTCATGGCT 1486

Qy   1834 CTCCAGCGGGGAGTGTTTCGCTGCTGTGATTACAGAGTCACCGGAGCACACCGTTCTGTGT 1893
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1485 CTCCAGCGGGGAGTGTTTCGCTGCTGTGATTACAGAGTCACCGGAGCACACCGTTCTGTGT 1426

Qy   1894 TACTGCTGTGGCCTGGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCT 1953
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1425 TACTGCTGTGGCCTGGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCT 1366

Qy   1954 GCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCTGACTGCTACTGCGGGGCCGTAACTGC 2013
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1365 GCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCTGACTGCTACTGCGGGGCCGTAACTGC 1306

Qy   2014 CGCACTCAGGTGAAAGCTCACCACGCCCATGAGAATTCAATCATATCTGTGAACAGACAAGG 2073
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1305 CGCACTCAGGTGAAAGCTCACCACGCCCATGAGAATTCAATCATATCTGTGAACAGACAAGG 1246

Qy   2074 TTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGCTTTCACAGCCCCCTGAGGAGGG 2133
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1245 TTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGCTTTCACAGCCCCCTGAGGAGGG 1186

Qy   2134 TGTTTTTAAAATACAGAGACAAGCACGTCAAGTGTTTCACAGCGTGTTTCACAGCCCCTCGA 2193
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1185 TGTTTTTAAAATACAGAGACAAGCACGTCAAGTGTTTCACAGCCCCCTGAGGGGACGGG 1126

Qy   2194 ACGCAGGGTCTCCGACAGTGCTCTCGGGGTGACTCTTCTGTGGGGAGCTTTTTACCCTCTCGA 2253
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1125 ACGCAGGGTCTCCACAGTGCTCTCGGGGTGACTCTTCTGTGGGAGC--TTTACCCTCTCGA 1068

Qy   2254 GTGAGACCCTCCCAGAGCCCGGGGGGGCCGCGCAGCCGCAGCCCCTCGGTGAGCGCTGGGCAG 2313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1067 GTGAGACCCTCCCCAGAGCCCCCGGGGGCCGCAGCCGCAGCCCCTCGGTGAGCGCTGGGCAG 1008

Qy   2314 GGCTCGTGTGTGGCATCAGCAGCAGCAGCAGCAGAAGAAGCCTTTCTGTAACATGGCGGCCGTCCCCGCC 2373
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1007 GGCTCGTGTGTGGCATCAGCAGCAGCAGCAGAAGAAGCCTTTCTGTAACATGGCGGCCGTCCCCGCC 948

Qy   2374 GAGAGGGGCAGTTTGCTCTTTGTACATTTCCGAAACTACAGTTAAAGCAGAAGTCTG 2433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   947 GAGAGGGGCAGTTTGCTCTTTTGTACATTTTCGAAACTACAGTTAAAGCAGAAGTCTG 888

Qy   2434 TTTTCAGGAAAAGTTTCAAGGGAGGAGAAGGGCAAGTTATTCAAAAACATTGTTTCAGGAGAA 2493
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   887 TTTTCAGGAAAAGTTTCAAGGGAGGAGAAGGGCAAGTTATTCAAAAACATTGTTTCAGGAGAA 828

Qy   2494 GGGAGCATAAGTTACAGCTACAGGACGTACCAGGATGCACCC 2553
          ||||||||||||||||||||||||||||||||||||||||||
Db   827 GGGAGCATAAGTTACAGCTACAGGACGTACCAGGATGCACCA 768

Qy   2554 GCATTTTATCTATTTTTATTTTAATAGGTTTGTGCTTATCTTCTAATAAGATTTAAAT 2613
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   767 GCATTTTATCTATTTTTATTTTTAATAGGTTTGGTTGCTTATCTTCTAATAAGATTTAAAT 708

Qy   2614 GTCACAAACTGTAGCACAAATATAATATTATATTTATATAATTTACAAAATTGAC 2661
          |||||||||||||||||||||||||||||||||||||||||
Db   707 GTCACAAACTGTAGCACAAATATAATATTAATTTATAATTTACAAAATTGAC 660

RESULT 7
AAS25843
ID   AAS25843 standard; cDNA; 1311 BP.
XX
AC   AAS25843;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human cDNA encoding a novel secreted protein, Seq ID 22.

```
XX
KW   Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   11-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0209963P.
PR   26-JUL-2000; 2000US-0209964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.

PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
```

```
Db     683  GCCAGGACCTGCTGCACGACTGCGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCGG   742
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy    1070  CTTGCTACTCGGGCTGGATGGAGCGCTCGTCCGTTGTCCTGGTCCTGCCGTGTCCCGTGG  1129
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     743  CTTGCTACTCGGGCTGGATGGAGCGCTCGTCCGTTGTCCTGGTCCTGCCGTGTCCCGTGG   802

Qy    1130  AGCGGATCTGTAAAACCACATCCTCAACAACCTCGTGGAAGCATACCTCATCCAGCATC   1189
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     803  AGCGGATCTGTAAAACCACATCCTCAACAACCTCGTGGAAGCATACCTCATCCAGCATC    862

Qy    1190  CAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTC  1249
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     863  CAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTC   922

Qy    1250  AAGACATGCTGCAGCCCAAAGTCAGGCGGGTCTTTTCTGATGAAGAAGGGAGTTCAGAGG  1309
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     923  AAGACATGCTGCAGCCCAAAGTCAGGCGGGTCTTTTCTGATGAAGAAGGGAGTTCAGAGG   982

Qy    1310  ACCTGCTGGAGCTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTCG  1369
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     983  ACCTGCTGGAGCTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTCG  1042

Qy    1370  TGTGCCGCAGTGTCCTGAGTACAGAAGGCAGGCGGCGCAGCCTCCCCACTGCCCAGCAC  1429
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1043  TGTGCCGCAGTGTCCTGAGTACAGAAGGCAGGCGGCGCAGCCTCCCCACTGCCCAGCAC  1102

Qy    1430  CCGAGGGCGGAGCCAGGAGCCCCACAGGCCCTGGGGATGCACCCT  1474
            |||||||||||||||||||||||||||||||||||||||||||||
Db    1103  CCGAGGGGCCGAGCCAGGAGCCCCACAGCCCTGGGGGATGCACCCT  1147
```

RESULT 8

```
Query Match         28.6%;  Score 765;  DB 4;  Length 1311;
Best Local Similarity  100.0%;  Pred. No. 2.2e-273;
Matches   765;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    710  GGTCTGGGGGTGGTGGTGCCATCTCCCCTAAAGGAAGGTGGTCCCCTGTGTGGCAAGTGATGAAG   769
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    383  GGTCTGGGGGGTGGTGGCATCTCCCCTAAAGGAAGGTGGTCCCCTGTGTGGCAAGTGATGAAG    442

Qy    770  TCTCCAGCTTGCCTCAGCTCTCCCAGACAGAAAGACTGCGTCCTTTTCGTCGTTGGAAC   829
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    443  TCTCCAGCTTGCCTCAGCTCTCCAGACAGAAAGACTGCGTCCTTTTCGTCGTTGGAAC    502

Qy    830  CCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAATGAGAGGAGATGGGGACCTTG   889
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    503  CCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAATGAGAGGAGATGGGGACCTTG    562

Qy    890  ACCTGAACGGGCAGTTGTTGGTCGCACAACCGGTAGAAATGCCCAAACCGTCCACGAGG   949
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    563  ACCTGAACGGGCAGTTGTTGGTCGCACAACCGGTAGAAATGCCCAAACCGTCCACGAGG    622

Qy    950  ACGTCAGAGCAGCGGCTGGGAAGCCAGACAAGATGGAGGAGACGCTGACATGCATCATCT  1009
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    623  ACGTCAGAGCAGCGGCTGGGAAGCCAGACAAGATGGAGGAGACGCTGACATGCATCATCT   682

Qy   1010  GCCAGGACCTGCTGCACGACTGCGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCGG  1069
```

```
PR    14-AUG-2000;  2000US-0225447P.
PR    14-AUG-2000;  2000US-0225757P.
PR    14-AUG-2000;  2000US-0225758P.
PR    22-AUG-2000;  2000US-0226868P.
PR    30-AUG-2000;  2000US-0228924P.
PR    01-SEP-2000;  2000US-0229287P.
PR    01-SEP-2000;  2000US-0229343P.
PR    01-SEP-2000;  2000US-0229344P.
PR    01-SEP-2000;  2000US-0229345P.
PR    05-SEP-2000;  2000US-0229509P.
PR    05-SEP-2000;  2000US-0229513P.
PR    08-SEP-2000;  2000US-0231413P.
PR    21-SEP-2000;  2000US-0234223P.
PR    21-SEP-2000;  2000US-0234274P.
PR    25-SEP-2000;  2000US-0234997P.
PR    27-SEP-2000;  2000US-0235834P.
PR    29-SEP-2000;  2000US-0236327P.
PR    29-SEP-2000;  2000US-0236367P.
PR    29-SEP-2000;  2000US-0236368P.
PR    29-SEP-2000;  2000US-0236369P.
PR    29-SEP-2000;  2000US-0236370P.
PR    02-OCT-2000;  2000US-0236802P.
PR    02-OCT-2000;  2000US-0237037P.
PR    02-OCT-2000;  2000US-0237038P.
PR    02-OCT-2000;  2000US-0237039P.
PR    02-OCT-2000;  2000US-0237040P.
PR    13-OCT-2000;  2000US-0239935P.
PR    20-OCT-2000;  2000US-0240960P.
PR    20-OCT-2000;  2000US-0241785P.
PR    20-OCT-2000;  2000US-0241809P.
PR    20-OCT-2000;  2000US-0244617P.
PR    17-NOV-2000;  2000US-0049299P.
PR    08-DEC-2000;  2000US-0251856P.
PR    08-DEC-2000;  2000US-0251868P.
PR    08-DEC-2000;  2000US-0251869P.
XX
PA    (ROSE/)  ROSEN C A.
PA    (RUBE/)  RUBEN S M.
PA    (BARA/)  BARASH S C.
XX
PI    Rosen CA,  Ruben SM,  Barash SC;
XX
DR    WPI; 2003-147444/14.
DR    P-PSDB; ABU54924.
XX
PT    New polypeptides and nucleic acids, useful in gene therapy for treating,
PT    inhibiting or preventing e.g. neural, immune system, muscular,
PT    respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT    renal disorders.
XX
PS    Claim 1; SEQ ID NO 22; 402pp; English.
XX
CC    The invention relates to human novel polypeptides and their associated
CC    polynucleotides. The polypeptides and polynucleotides are useful in gene
CC    therapy for treating, inhibiting or preventing neural disorders, immune
CC    system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC    and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC    nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC    gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC    (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC    heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC    renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC    leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC    appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC    related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC    infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC    human novel polynucleotides of the invention
XX
SQ    Sequence 1311 BP; 306 A; 359 C; 399 G; 239 T; 0 U; 8 Other;

Query Match          28.6%;  Score 765;  DB 7;  Length 1311;
Best Local Similarity  100.0%;  Pred. No. 2.2e-273;
Matches  765;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Qy   710  GGTCTGGGGTGGTGCATCCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAAG  769
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   383  GGTCTGGGGTGGTGCATCTCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAAG  442

Qy   770  TCTCCAGCTTTGCCTCAGCTCTCCAGACAGAAGACTGCGTCCTTTTCGTCGTTGGAAC  829
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   443  TCTCCAGCTTTGCCTCAGCTCTCCAGACAGAAGACTGCGTCCTTTTCGTCGTTGGAAC  502

Qy   830  CCCAGGATCAGGAGGAGATTTGGAGCCCGTGAAGAGAAAATGAGAGACGAGATGGGGACCTTG  889
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   503  CCCAGGATCAGGAGGAGATTTGGAGCCCGTGAAGAGAAAATGAGAGACGAGATGGGGACCTTG  562

Qy   890  ACCTGAACGGGCAGTTGTTGTCGCACAACCGGTAGAAATGCCCAAACCGTCCACGAGG  949
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   563  ACCTGAACGGGCAGTTGTTGTCGCACAACCGGTAGAAATGCCCAAACCGTCCACGAGG  622

Qy   950  AGTCGAGACGCGGCTGGAAGCCAGACAAGATGGAGGAGACCGCTGACATGCATCATCT  1009
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   623  AGTCAGAGACGCGGCTGGAAGCCAGACAAGATGGAGGAGACCGCTGACATGCATCATCT  682

Qy  1010  GCCAGGACCTGCTGACGACTGCGTGAGTTGCAGCCCTGTCCCTGTCCTCACCTGCTGCCCGTG  1069
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   683  GCCAGGACCTGCTGACGACTGCGTGAGTTGCAGCCCTGTCCCTGTCCTCACCTGCTGCCCGTG  742

Qy  1070  CTTGCTACTCGGGCTGGATGGAGCGCTCGTCCCTGTCCTCAACAACCTCGTGAAGCATACCTCATCCAGCATC  1129
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   743  CTTGCTACTCGGGCTGGATGGAGCGCTCGTCCCTGTCCTCAACAACCTCGTGAAGCATACCTCATCCAGCATC  802

Qy  1130  AGCGGATCTGTAAAAACCACATCCTCAACAACCTCGTGAAGAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTC  1189
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   803  AGCGGATCTGTAAAAACCACATCCTCAACAACCTCGTGAAGAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTC  862

Qy  1190  CAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTC  1249
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   863  CAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTC  922

Qy  1250  AAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTTCTGATGAACAAGGGAGTTCAGAGG  1309
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   923  AAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTTCTGATGAACAAGGGAGTTCAGAGG  982

Qy  1310  ACCTGCTGGAGCTGTCAGAGCTGTGACAGTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTCG  1369
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   983  ACCTGCTGGAGCTGTCAGAGCTGTGACAGTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTCG  1042

Qy  1370  TGTGCCGGCAGTGTCCTGAGTACAGAGAGGCAGGGCGCGCAGCCTCCCCACTGCCCAGCAC  1429
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1043  TGTGCCGGCAGTGTCCTGAGTACAGAGAGGCAGGGCGCGCAGCCTCCCCACTGCCCAGCAC  1102

Qy  1430  CCGAGGGCGAGCCAGGAGCCCCACAGGCCCTGGGGGATGCACCCT  1474
               ||||||||||||||||||||||||||||||||||||||||||||
Db  1103  CCGAGGGGCGAGCCAGGAGCCCCACAGGCCCCACAGGCCCTGGGGGATGCACCCT  1147
```

```
XX
PF    21-MAR-2002; 2002EP-00006586.
XX
PR    14-SEP-2001; 2001JP-00328381.
PR    24-JAN-2002; 2002US-0350435P.
XX
PA    (HELI-) HELIX RES INST.
PA    (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI    Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI    Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI    Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR    WPI; 2003-395539/38.
DR    P-PSDB; ADA54231.
XX
PT    New polynucleotides encoding full-length polypeptides, e.g. secretory
PT    and/or membrane proteins, useful for developing medicines for diseases in
PT    which the gene is involved, or as target molecules for gene therapy.
XX
PS    Claim 1; SEQ ID NO 160; 205pp; English.
XX
CC    The present invention relates to novel human secretory or membrane
CC    proteins (ADA54072-ADA55710) and their coding sequences (ADA2433-
CC    ADA54071). The coding sequences are useful in the gene therapy of
CC    diseases caused by abnormalities of the proteins, e.g. cancer,
CC    inflammatory diseases, osteoporosis or neurological disease.
XX
SQ    Sequence 2186 BP; 493 A; 555 C; 539 G; 599 T; 0 U; 0 Other;

Query Match           23.6%;  Score 633;  DB 7;  Length 2186;
Best Local Similarity 99.7%;  Pred. No. 1.4e-224;
Matches  863;  Conservative  0;  Mismatches  1;  Indels  2;  Gaps  1;

Qy     1796 TGACATGGAAAAACATGTTGACCGAGAGCCTCGTGGCCTCCAGCGGGGAGTGTTCTGC 1855
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      695 TGACATGGAAAAACATGTTGACCGAGAGCCTCGTGGCCTCCAGCGGGGAGTGTTCTGC 754

Qy     1856 TGTCTGATTACAGAGTCACGGAGACACCGTTCGTTCGTGTTACTGCTGTGGCCTGCGCAGCT 1915
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      755 TGTCTGATTACAGAGTCACGGAGACACCGTTCGTTCGTGTTACTGCTGTGGCCTGCGCAGCT 814

Qy     1916 TCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCG 1975
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      815 TCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCG 874

Qy     1976 TAACATCCCGTCCTGACTGCTACTGGGGCCGTAACTGCGCACTCAGGTGAAAGCTCACC 2035
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      875 TAACATCCCGTCCTGACTGCTACTGGGGCCGTAACTGCGCACTCAGGTGAAAGCTCACC 934

Qy     2036 ACGCCATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAG 2095
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      935 ACGCCATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAG 994

Qy     2096 GCCCTGAGCAGCTTTCAGCACTGGAGGTGAAGAGAGCGTGTTTTTAAAATACAGAGACAA 2155
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      995 GCCCTGAGCAGCTTTCAGCACTGGAGGTGAAGAGAGCGTGTTTTTAAAATACAGAGACAA 1054

Qy     2156 GCACGTCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGGACGCAGAGGGTCCGACAGGTGC 2215
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1055 ACACGTCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGGACGCAGAGGGTCCGACAGGTGC 1114

Qy     2216 TCTCGGGGTGACTCCTTCTGGGAGCTTTTTACCCTCTGAGTGAGACCCTCCCCAGAGCCCC 2275
            |||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db     1115 TCTGGGGTGACTCCTTCTGTGTGGAGC--TTTACCCTCTGAGTGAGACCCTCCCCAGAGCCCC 1172

Qy     2276 GGGGGCCGCAGCCCGGCCCCTCCGTGAGCCTGGCAGGGCTCGTGGTGGCATCAGCAGC 2335
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1173 GGGGGCCCCAGCCCGGCCCCTCCGTGAGCCTGGCAGGGCTCGTGGTGGCATCAGCAGC 1232

Qy     2336 AGAGACGAAGCCTTTCTGTAACATGCGGCCGTCCGCCCAGAGGGGCAGTTTTGCTCTTT 2395
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1233 AGAGACGAAGCCTTTCTGTAACATGCGGCCGTCCGCCCAGAGGGGCAGTTTTGCTCTTT 1292
```

```
Qy     2396 TGTACATTTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAAGTTTCAAGGG 2455
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1293 TGTACATTTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAAGTTTCAAGGG 1352

Qy     2456 AGAAGGGCAAGTTATCAAAAACATTGTTTCAGGAGAAGGAGCATAAGTTTACAGGCCTA 2515
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1353 AGAAGGGCAAGTTATCAAAAACATTGTTTCAGGAGAAGGAGCATAAGTTTACAGCCTA 1412

Qy     2516 CAGGACCTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTTTTATTT 2575
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1413 CAGGACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTTTTTATTT 1472

Qy     2576 TAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATA 2635
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1473 TAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATA 1532

Qy     2636 ATATAATTATAATTTACAAATTGAC 2661
            |||||||||||||||||||||||||
Db     1533 ATATAATTATAATTTACAAATTGAC 1558

RESULT 10
AAH06828
ID    AAH06828 standard; cDNA; 816 BP.
XX
AC    AAH06828;
XX
DT    26-JUN-2001  (first entry)
XX
DE    Human cDNA clone (5'-primer) SEQ ID NO:3663.
XX
KW    Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS    Homo sapiens.
XX
PN    EP1074617-A2.
XX
PD    07-FEB-2001.
XX
PF    28-JUL-2000; 2000EP-00116126.
XX
PR    29-JUL-1999;     99JP-00248036.
PR    27-AUG-1999;     99JP-00300253.
PR    11-JAN-2000;     2000JP-00118776.
PR    02-MAY-2000;     2000JP-00183767.
PR    09-JUN-2000;     2000JP-00241899.
XX
PA    (HELI-) HELIX RES INST.
XX
PI    Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI    Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR    WPI; 2001-318749/34.
XX
PT    Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT    length cDNAs defined in the specification, and for the detection and/or
PT    diagnosis of the abnormality of the proteins encoded by the full-length
PT    cDNAs.
XX
PS    Claim 1; SEQ ID NO 3663; 2537pp + Sequence Listing; English.
XX
CC    The present invention describes primer sets for synthesising 5602 full-
CC    length cDNAs defined in the specification. Where a primer set comprises:
CC    (a) an oligo-dT primer and an oligonucleotide complementary to the
CC    complementary strand of a polynucleotide which comprises one of the 5602
CC    nucleotide sequences defined in the specification, where the
CC    oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC    of an oligonucleotide comprising a sequence complementary to the
CC    complementary strand of a polynucleotide which comprises a 5'-end
CC    sequence and an oligonucleotide comprising a sequence complementary to a
CC    polynucleotide which comprises a 3'-end sequence, where the
CC    oligonucleotide comprises at least 15 nucleotides and the combination of
```

```
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the protein encoded by the full-length
PT   cDNAs.
XX
PS   Claim 1; SEQ ID NO 3681; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 824 BP; 200 A; 203 C; 258 G; 161 T; 0 U; 2 Other;

Query Match             17.5%;  Score 468;  DB 4;  Length 824;
Best Local Similarity  100.0%;  Pred. No. 1.9e-163;
Matches  468;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy         27 CTCTTGACAGCGGCCGGCGGCGGCCCGAGCAGTCGGCGGTTCGGCGTTCGGCGGGGATGTGAAT 86
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTCTTGACAGCGGCCGGCGGCGGCCCGAGCAGTCGGCGGTTCGGCGTTCGGCGGGGATGTGAAT 60

Qy         87 CCCGATGGAGCGGAGAGCGAGGTTGCGACCTTTCCTTCCCCAGCAATAAAACTGGTCTCTGGAGA 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCCGATGGAGCGGAGAGCGAGGTTGCGACCTTTCCTTCCCCAGCAATAAAACTGGTCTCTGGAGA 120

Qy        147 CCTGCGTCTCGGGCGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAAACTGGTGGAC 206
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTGCGTCTCGGGCGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAAACTGGTGGAC 180

Qy        207 CATCGGGCGGAGACGAGGTTGCGACCTGGACCTTTCCTTCCCCAGCAATAAACTGGTGCTCTGGAGA 266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATCGGGCGGAGACGAGGTTGCGACCTGGACCTTTCCTTCCCCAGCAATAAACTGGTGCTCTGGAGA 240

Qy        267 TCACTGTAGAATTGTAGTGGATGAAAATCAGGTCAGGTGACACTGGAAGATACCAGCAC 326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCACTGTAGAATTGTAGTGGATGAAAATCAGGTCAGGTGACACTGGAAGATACCAGCAC 300

Qy        327 CAGTGGAACAGTGATTAACAGCTGAAGGTTGTTAAGAAGCAGACATGCCCCTTTACAGAC 386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CAGTGGAACAGTGATTAACAGCTGAAGGTTGTTAAGAAGCAGACATGCCCCTTTACAGAC 360

Qy        387 TGGGGATGGTCATCTACTTGGTGTACAGGAAGAATGAACCGGAACACAACGTGGCATACCT 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 816 BP; 182 A; 216 C; 253 G; 162 T; 0 U; 3 Other;

Query Match             17.8%;  Score 478;  DB 4;  Length 816;
Best Local Similarity  100.0%;  Pred. No. 3.8e-167;
Matches  478;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy         17 GCCGCAATGTCTCTTGACAGCGGCGGCGGCGGCCCGAGCCGGTTCGGCGGTTCGGCGGGGCGG 76
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 GCCGCAATGTCTCTTGACAGCGGCGGCGGCGGCCCGAGCCGGTTCGGCGGTTCGGCGGGGCGG 64

Qy         77 GGATGTGAATCCCGATGGAGCGGAGAGCGAGGCAAGCAGTCGCCGCGCGCGCAGCCCT 136
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 GGATGTGAATCCCGATGGAGCGGAGAGCGAGGCAAGCAGTCGCCGCGCGCGCAGCCCT 124

Qy        137 GGGGACGGCTCCTGCGTCTGGGCGCGGAGGAGGGCCGAGCCGCACGTCCTCCTGAGGAAGC 196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 GGGGACGGCGTCCTGCGTCTGGGCGCGGAGGAGGGCCGAGCCGCACGTCCTCCTGAGGAAGC 184

Qy        197 GGGAGTGGACCATCGGGCGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGG 256
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 GGGAGTGGACCATCGGGCGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGG 244

Qy        257 TCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGTCAGGTGACACTGGAAG 316
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 TCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGTCAGGTGACACTGGAAG 304

Qy        317 ATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGACATGCC 376
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 ATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGACATGCC 364

Qy        377 CTTTACAGACTGGGGATGTCATCATCTACTTGGTGTACAGAAGAATGAACCGGAACACACG 436
              ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 CTTTACAGACTGGGGATGTCATCATCTACTTGGTGTACAGGAAGAATGAACCGGAACACACG 424

Qy        437 TGGCATACCTCTCATGAATCTTTAAGTGAAAAGCAAGGCATGACACAACAAGAATCCTTTGA 494
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 TGGCATACCTCTCATGAATCTTTAAGTGAAAAGCAAGGCATGACACAACAAGAATCCTTTGA 482

RESULT 11
AAH06846
ID   AAH06846 standard; cDNA; 824 BP.
XX
AC   AAH06846;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:3681.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000BP-00116126.
XX
PR   29-JUL-1999;  99JP-00248036.
PR   27-AUG-1999;  99JP-00300253.
```

```
Db      361  TGGGGATGTCATCTACTTGGTGTACAGGAAGAATGACCGGAACACAACGTGGCATACCT 420

Qy      447  CTATGAATCTTTAAGTGAAAAGCAAGGCATGACACAAGAATCCTTTGA 494
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db      421  CTATGAATCTTTAAGTGAAAAGCAAGGCATGACACAAGAATCCTTTGA 468


RESULT 12
AAS26304
ID    AAS26304 standard; cDNA; 693 BP.
XX
AC    AAS26304;
XX
DT    07-NOV-2001  (first entry)
XX
DE    Human cDNA encoding a novel secreted protein, Seq ID 483.
XX
KW    Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;
KW    cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW    antibacterial; virucide; fungicide; opthalmological; vulnerary;
KW    secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW    cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW    cerebral ischaemia; angiogenesis; nervous system disorder;
KW    Alzheimer's disease; infection; ocular disorder; corneal infection;
KW    wound healing; epithelial cell proliferation; skin ageing; food additive;
KW    preservative; antiproliferative.
XX
OS    Homo sapiens.
XX
PN    WO200155322-A2.
XX
PD    02-AUG-2001.
XX
PF    17-JAN-2001; 2001WO-US001341.
XX
PR    31-JAN-2000;  2000US-0179065P.
PR    04-FEB-2000;  2000US-0180628P.
PR    24-FEB-2000;  2000US-0184664P.
PR    02-MAR-2000;  2000US-0186350P.
PR    16-MAR-2000;  2000US-0189874P.
PR    17-MAR-2000;  2000US-0190076P.
PR    18-APR-2000;  2000US-0198123P.
PR    19-MAY-2000;  2000US-0205515P.
PR    07-JUN-2000;  2000US-0209467P.
PR    28-JUN-2000;  2000US-0214886P.
PR    30-JUN-2000;  2000US-0215135P.
PR    07-JUL-2000;  2000US-0216647P.
PR    07-JUL-2000;  2000US-0216880P.
PR    11-JUL-2000;  2000US-0217487P.
PR    11-JUL-2000;  2000US-0217496P.
PR    14-JUL-2000;  2000US-0218290P.
PR    26-JUL-2000;  2000US-0220963P.
PR    26-JUL-2000;  2000US-0220964P.
PR    14-AUG-2000;  2000US-0224518P.
PR    14-AUG-2000;  2000US-0224519P.
PR    14-AUG-2000;  2000US-0225267P.
PR    14-AUG-2000;  2000US-0225213P.
PR    14-AUG-2000;  2000US-0225214P.
PR    14-AUG-2000;  2000US-0225266P.
PR    14-AUG-2000;  2000US-0225267P.
PR    14-AUG-2000;  2000US-0225268P.
PR    14-AUG-2000;  2000US-0225270P.
PR    14-AUG-2000;  2000US-0225447P.
PR    14-AUG-2000;  2000US-0225757P.
PR    14-AUG-2000;  2000US-0225758P.
PR    14-AUG-2000;  2000US-0225759P.
PR    18-AUG-2000;  2000US-0226279P.
PR    22-AUG-2000;  2000US-0226681P.
PR    22-AUG-2000;  2000US-0226868P.
PR    22-AUG-2000;  2000US-0227182P.
PR    23-AUG-2000;  2000US-0227009P.
PR    30-AUG-2000;  2000US-0228924P.
PR    01-SEP-2000;  2000US-0229287P.
PR    01-SEP-2000;  2000US-0229343P.
PR    01-SEP-2000;  2000US-0229344P.
PR    01-SEP-2000;  2000US-0229345P.
PR    05-SEP-2000;  2000US-0229509P.
PR    05-SEP-2000;  2000US-0229513P.
PR    06-SEP-2000;  2000US-0230437P.
PR    06-SEP-2000;  2000US-0230438P.
PR    08-SEP-2000;  2000US-0231242P.
PR    08-SEP-2000;  2000US-0231243P.
PR    08-SEP-2000;  2000US-0231244P.
PR    08-SEP-2000;  2000US-0231413P.
PR    08-SEP-2000;  2000US-0231414P.
PR    08-SEP-2000;  2000US-0232080P.
PR    08-SEP-2000;  2000US-0232081P.
PR    12-SEP-2000;  2000US-0231968P.
PR    14-SEP-2000;  2000US-0232398P.
PR    14-SEP-2000;  2000US-0232399P.
PR    14-SEP-2000;  2000US-0232400P.
PR    14-SEP-2000;  2000US-0232401P.
PR    14-SEP-2000;  2000US-0233063P.
PR    14-SEP-2000;  2000US-0233064P.
PR    14-SEP-2000;  2000US-0233065P.
PR    21-SEP-2000;  2000US-0234223P.
PR    21-SEP-2000;  2000US-0234274P.
PR    25-SEP-2000;  2000US-0234997P.
PR    25-SEP-2000;  2000US-0234998P.
PR    26-SEP-2000;  2000US-0235484P.
PR    27-SEP-2000;  2000US-0235834P.
PR    27-SEP-2000;  2000US-0235836P.
PR    29-SEP-2000;  2000US-0236327P.
PR    29-SEP-2000;  2000US-0236367P.
PR    29-SEP-2000;  2000US-0236368P.
PR    29-SEP-2000;  2000US-0236369P.
PR    29-SEP-2000;  2000US-0236370P.
PR    02-OCT-2000;  2000US-0236802P.
PR    02-OCT-2000;  2000US-0237037P.
PR    02-OCT-2000;  2000US-0237038P.
PR    02-OCT-2000;  2000US-0237039P.
PR    02-OCT-2000;  2000US-0237040P.
PR    13-OCT-2000;  2000US-0239935P.
PR    13-OCT-2000;  2000US-0239937P.
PR    20-OCT-2000;  2000US-0240060P.
PR    20-OCT-2000;  2000US-0241221P.
PR    20-OCT-2000;  2000US-0241785P.
PR    20-OCT-2000;  2000US-0241786P.
PR    20-OCT-2000;  2000US-0241787P.
PR    20-OCT-2000;  2000US-0241808P.
PR    20-OCT-2000;  2000US-0241809P.
PR    20-OCT-2000;  2000US-0241826P.
PR    01-NOV-2000;  2000US-0244617P.
PR    08-NOV-2000;  2000US-0246474P.
PR    08-NOV-2000;  2000US-0246475P.
PR    08-NOV-2000;  2000US-0246476P.
PR    08-NOV-2000;  2000US-0246477P.
PR    08-NOV-2000;  2000US-0246478P.
PR    08-NOV-2000;  2000US-0246523P.
PR    08-NOV-2000;  2000US-0246524P.
PR    08-NOV-2000;  2000US-0246525P.
PR    08-NOV-2000;  2000US-0246526P.
PR    08-NOV-2000;  2000US-0246527P.
PR    08-NOV-2000;  2000US-0246528P.
PR    08-NOV-2000;  2000US-0246532P.
PR    08-NOV-2000;  2000US-0246609P.
PR    08-NOV-2000;  2000US-0246610P.
PR    08-NOV-2000;  2000US-0246611P.
PR    17-NOV-2000;  2000US-0246613P.
PR    17-NOV-2000;  2000US-0249207P.
PR    17-NOV-2000;  2000US-0249208P.
PR    17-NOV-2000;  2000US-0249209P.
PR    17-NOV-2000;  2000US-0249210P.
PR    17-NOV-2000;  2000US-0249211P.
PR    17-NOV-2000;  2000US-0249212P.
PR    17-NOV-2000;  2000US-0249213P.
```

```
Db      74  GTGTCCTACCTGCCGCTGTCCCGTGGAGCGGATCTGTAAAAACCACATCCTCAACAACCT  133
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1164  CGTGGAAGCATACCTCATCCAGCATCCAGACAAGAGTCGCAGTGAAGAGATGTGCAAAG  1223
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     134  CGTGGAAGCATACCTCATCCAGCATCCAGACAAGAGTCGCAGTGAAGAGATGTGCAAAG  193

Qy    1224  TATGGATGCCAGGAATAAAATCACTCAAGACATGCTGAGCTGTCAGCCCAAAGTCAGGCGGTCTTT  1283
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     194  TATGGATGCCAGGAATAAAATCACTCAAGACATGCTGAGCTGTCAGCCCAAAGTCAGGCGGTCTTT  253

Qy    1284  TTCTGATGAAGAAGGGAGTTCAGAGACCTGCTGGAGCTGTCGTCAGACGTTGACAGTGAGTC  1343
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     254  TTCTGATGAAGAAGGGAGTTCAGAGACCTGCTGGAGCTGTCGTCAGACGTTGACAGTGAGTC  313

Qy    1344  CTCAGACATTAGCCAGCCATACGTCGTGTGCGCAGTGTCCTGAGTACAGAAGGCAGGC  1403
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     314  CTCAGACATTAGCCAGCCATACGTCGTGTGCGCAGTGTCCTGAGTACAGAAGGCAGGC  373

Qy    1404  GGCGCAGCCTCCCACTGCCAGCACCCGAGGGCGAGCCAGGAGCCCCACAGGCCCTGGG  1463
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     374  GGCGCAGCCTCCCACTGCCAGCACCCGAGGGCGAGCCAGGAGCCCCACAGGCCCTGGG  433

Qy    1464  GGATGCACCCTCCACGTCCGTCAGCCTGACGACAG  1498
            ||||||||||||||||||||||||||||||||||
Db     434  GGATGCACCCTCCACGTCCGTCAGCCTGACGACAG  468
```

RESULT 13
ABX73645
ID   ABX73645 standard; DNA; 693 BP.
XX
AC   ABX73645;
XX
DT   18-MAR-2003 (first entry)
XX
DE   Human novel polynucleotide #473.
XX
KW   Human; gene; ds; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225575P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   P-PSDB; AAU16317.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 1; SEQ ID NO 483; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence encodes a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed

Query Match   17.0%; Score 455; DB 4; Length 693;
Best Local Similarity   100.0%; Pred. No. 1.3e-158;
Matches 455; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1044 GCCCTGCATGCACACGTTCTGCGCCGCTTGCTACTCGGGCTGGATGGAGGCCTCGTCCCT 1103
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14 GCCCTGCATGCACACGTTCTGCGCCGCTTGCTACTCGGGCTGGATGGAGGCCTCGTCCCT   73

Qy   1104 GTGTCCTACCTGCCGCTGTCCCGTGGAGCGGATCTGTAAAAACCACATCCTCAACAACCT 1163
```

```
Qy  1104  GTGTCCTACCTGCGCGTCCCGTGGAGCGGATCTGTAAAACCACATCCTCAACACCT  1163
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    74  GTGTCCTACCTGCGCGTCCCGTGGACGCGGATCTGTAAAAACCACATCCTCAACACCT  133

Qy  1164  CGTGGAAGCATACCTCATCCAGCATCCAGACAAGAGTCGCAGTGAAGAGATGTGCAAAG  1223
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   134  CGTGGAAGCATACCTCATCCAGCATCCAGACAAGAGTCGCAGTGAAGAGATGTGCAAAG  193

Qy  1224  TATGGATGCCAGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTT  1283
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   194  TATGGATGCCAGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTT  253

Qy  1284  TTCTGATGAAGAAGGGAGTTCAGAGACCTGCTGGAGCTGTCAGACGTTGACAGTGAGTC  1343
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   254  TTCTGATGAAGAAGGAGTTCAGAGACCTGCTGGAGCTGTCAGACGTTGACAGTGAGTC  313

Qy  1344  CTAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGC  1403
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   314  CTCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAAGGAGGC  373

Qy  1404  GGGCGAGCCTCCCACTGCCCAGCACCCGAGGGCGAGCCAGGAGCCCCAGCCCTGG  1463
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   374  GGGCGAGCCTCCCACTGCCCAGCACCCGAGGGCGAGCCAGGAGCCCCAGCCCTGG  433

Qy  1464  GGATGCACCCTCCACGTCCGTCAGCCTGACGACAG  1498
                |||||||||||||||||||||||||||||||||
Db   434  GGATGCACCCTCCACGTCGTCAGCCTGACGACAG  468
```

```
RESULT 14
AAH11859/c
ID   AAH11859 standard; cDNA; 518 BP.
XX
AC   AAH11859;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (3'-primer) SEQ ID NO:8694.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;    99JP-00248036.
PR   27-AUG-1999;    99JP-00300253.
PR   11-JAN-2000;    2000JP-00118776.
PR   02-MAY-2000;    2000JP-0018376?.
PR   09-JUN-2000;    2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyana T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 3; SEQ ID NO 8694; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
```

```
PR   30-AUG-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0229287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   01-SEP-2000;   2000US-0229345P.
PR   05-SEP-2000;   2000US-0229509P.
PR   05-SEP-2000;   2000US-0229513P.
PR   08-SEP-2000;   2000US-0231413P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   27-SEP-2000;   2000US-0235834P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241809P.
PR   01-NOV-2000;   2000US-0244617P.
PR   17-NOV-2000;   2000US-0249299P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA, Ruben SM, Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   P-PSDB; ABU55385.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 1; SEQ ID NO 483; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC   human novel polynucleotides of the invention
XX
SQ   Sequence 693 BP; 145 A; 214 C; 197 G; 123 T; 0 U; 14 Other;

Query Match    17.0%;  Score 455;  DB 7;  Length 693;
Best Local Similarity  100.0%;  Pred. No. 1.3e-158;
Matches  455;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  1044  GCCCTGCATGCACACGTTCTGCGCGGCTTGCTACTCGGGGTGGATGGAGCGGCTCGTCCCT  1103
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14  GCCCTGCATGCACACGTTCTGCGCGGCTTGCTACTCGGGGTGGATGGAGCGGCTCGTCCCT  73
```

Query Match   15.9%;   Score 426; DB 4; Length 518;
Best Local Similarity 99.8%;   Pred. No. 7.2e-148;
Matches 476; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy  2185 AGGGAAGGGACGCAGGGGTCTCCGACAGGTGCTTCTGGGGTGACTCTTCGTGTGGAGCTTTTT 2244
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   503 AGGGAAGGGACGCGCAGGGTCTTCCGACAGGTGCTCGGGGTGACTCTTCGTGTGGAGCTTTTT 444

Qy  2245 ACCCTCTGAGTGAGACCCTCCCCAGAGCCCCGGGGGCGCAGCCCGCCCTCCTGGTGAGC 2304
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   443 ACCCTCTGAGTGAGACCCTCCCCAGAGCCCCGGGGGCCCGCAGCCCGCCCTCCTGGTGAGC 384

Qy  2305 GCTGGGCAGGGCTCCTGGTGGCATCAGCAGCAGAGACGAAGCCTTCTGTAACATGCGGC 2364
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   383 GCTGGGCAGGGCTCGTGGTGGCATCAGCAGCAGAGACGAAGCCTTCTGTAACATGCCGGC 324

Qy  2365 CGTCCCGCCGAGAGGGGCAGTTTGCTCTTTTGTACATTTCCGAAACTACAGTTAAAGC 2424
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   323 CGTCCCGCCGAGAGGGGCAGTTTGCTCTTTTGTACATTTCCGAAACTACAGTTAAAGC 264

Qy  2425 AGAAGTCTGTTTTCAGGAAAAGTTTCAAGGGAGGAAGGGCAAGTTTATCAAAAACATTGTT 2484
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   263 AAAAGTCTGTTTTCAGGAAAAGTTTCAAGGGAGGAAGGGCAAGTTTATCAAAAACATTGTT 204

Qy  2485 TCAGGAGAAGGGAGCATAAGTTTACAGCCTACACAGGACGTACACAATATCCTGCTGCTGGG 2544
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   203 TCAGGAGAAGGGAGCATAAGTTTACAGCCCTACACAGGACGTACACAATATCCTGCTGCTGGG 144

Qy  2545 AAAACCACAGCATTTTATCATTTTATTTTAATTGGTGGCTTATCTTCTAATAA 2604
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   143 AAAACCACAGCATTTTATCTATTTTTTATTTTAATTGGTGGTTATCTTCTAATAA 84

Qy  2605 GATTTAAATGTCACAAACTGTAGCACAAATAATAATTATAATTTATAATTTACAAATTGAC 2661
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    83 GATTTAAATGTCACAAACTGTAGCACAAATAATAATTATAATTTATAATTTACAAATTGAC 27
```

RESULT 15
AAA44336
ID   AAA44336 standard; cDNA; 575 BP.
XX
AC   AAA44336;
XX
DT   21-AUG-2000 (first entry)
XX
DE   Human secreted expressed sequence tag SEQ ID NO:911.
XX
KW   Human; mouse; chicken; rat; secreted expressed sequence tag; sEST;
KW   expressed sequence tag; EST; probe; chemotactic; proliferative;
KW   immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW   thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;

Query Match   10.7%;   Score 286; DB 3; Length 575;
Best Local Similarity 100.0%;   Pred. No. 4.5e-96;
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   876 AGATGGGGACCTGACCTGAACCGGGCAGTTGTTGGTCGGCACAACCGGCTAGAAATGCCCA 935
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   285 AGATGGGGACCTTGACCTGAACCGGGCAGTTGTTGGTCGGCACAACCGGCTAGAAATGCCCA 344

Qy   936 AACCGTCCACGAGCAGTCAGAGCAGCGCGGCTCGGGAAGCCAGACAAGATGGGAGGAGACGCT 995
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   345 AACCGTCCACGAGCAGTCAGAGCAGCGCGGCTCGGGAAGCCAGACAAGATGGGAGGAGACGCT 404

Qy   996 GACATGCATCATCTGCCAGGACAGGACCTGCTGCTCGAGTTGCAGCCCTGCAGTGGCATGCA 1055
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    405  GACATGCATCATCTGCCAGGACCTGCTGCACGACTGCGTGAGTTTGCAGCCCTGCATGCA  464

Qy   1056  CACGTTCTGCGCGGCTTGCTACTCGGGGCTGGATGGAGCGCTCGTCCCTGTCCTACCTG  1115
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    465  CACGTTCTGCGCGGCTTGCTACTCGGGGCTGGATGGAGCGCTCGTCCCTGTCCTACCTG  524

Qy   1116  CCGCTGTCCCGTGGAGCGGATCGTAAAAACCACATCCTCAACAAC  1161
            |||||||||||||||||||||||||||||||||||||||||||||
Db    525  CCGCTGTCCCGTGGAGCGGATCGTAAAAACCACATCCTCAACAAC  570


Search completed: May 15, 2004, 00:35:33
Job time : 1488 secs
```

```
         Copyright    GenCore version 5.1.6
         Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    May 14, 2004, 23:43:56 ; Search time 214 Seconds
                                     (without alignments)
                                     6947.259 Million cell updates/sec

Title:          US-10-048-046-1
Perfect score:  2679
Sequence:       1 aagaattcggcacgaggccg........acaaaaaaaaaaaaaaaa 2679

Scoring table:  OLIGO_NUC
                Gapop 60.0 , Gapext 60.0

Searched:       682709 seqs, 277475446 residues

Word size :     0

Total number of hits satisfying chosen parameters:    1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :  Issued_Patents_NA:*
      1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
      2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
      3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
      4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
      5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
      6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	26	1.0	860	1	US-07-847-010-18	Sequence 18, Appl
	2	25	1.0	1085	4	US-09-522-714-15	Sequence 15, Appl
	3	24	0.9	3157	6	5198347-3	Patent No. 5198347
	4	23	0.9	443	4	US-09-621-976-17572	Sequence 17572, A
	5	23	0.9	1185	4	US-09-614-912-161	Sequence 161, App
	6	23	0.9	1347	3	US-09-412-102-1	Sequence 1, Appli
	7	23	0.9	1347	3	US-09-217-787-1	Sequence 1, Appli
	8	23	0.9	2405	1	US-08-484-101B-35	Sequence 35, Appl
	9	23	0.9	2405	1	US-08-484-101B-49	Sequence 49, Appl
	10	23	0.9	2405	3	US-08-714-524D-35	Sequence 35, Appl
	11	23	0.9	2405	3	US-08-714-524D-49	Sequence 49, Appl
	12	22	0.8	277	1	US-08-244-113-18	Sequence 18, Appl
c	13	22	0.8	424	4	US-09-280-116-159	Sequence 159, App
	14	22	0.8	786	4	US-09-525-258-172	Sequence 172, App
c	15	22	0.8	906	4	US-09-205-258-190	Sequence 190, App
	16	22	0.8	914	4	US-09-475-316A-24	Sequence 24, Appl
	17	22	0.8	914	4	US-09-704-640-24	Sequence 24, Appl
	18	22	0.8	1019	4	US-09-183-861-19	Sequence 19, Appl
	19	22	0.8	1019	4	US-09-022-765-19	Sequence 19, Appl
	20	22	0.8	1019	4	US-09-551-974A-19	Sequence 19, Appl
	21	22	0.8	1019	4	US-09-565-501A-19	Sequence 19, Appl
	22	22	0.8	1019	4	US-09-639-206A-19	Sequence 19, Appl
	23	22	0.8	1019	3	US-09-874-923-19	Sequence 19, Appl
	24	22	0.8	1173	4	US-09-232-200-50	Sequence 50, Appl
	25	22	0.8	1173	4	US-09-232-197-50	Sequence 50, Appl
	26	22	0.8	1173	4	US-09-232-201-50	Sequence 50, Appl
	27	22	0.8	1173	4	US-09-232-195-50	Sequence 50, Appl
	28	22	0.8	1287	4	US-09-064-411A-35	Sequence 35, Appl
	29	22	0.8	1404	4	US-08-979-608A-10	Sequence 10, Appl
	30	22	0.8	1404	4	US-09-517-849-10	Sequence 10, Appl
	31	22	0.8	1404	4	US-09-616-289-10	Sequence 10, Appl
	32	22	0.8	1557	4	US-09-347-650-15	Sequence 15, Appl
	33	22	0.8	2339	2	US-09-268-140-11	Sequence 11, Appl
	34	22	0.8	2422	1	US-07-867-106-5	Sequence 5, Appli
c	35	22	0.8	2505	3	US-09-268-140-1	Sequence 1, Appli
	36	22	0.8	2517	4	US-09-268-140-7	Sequence 7, Appli
	37	22	0.8	2533	4	US-09-604-605-1	Sequence 1, Appli
	38	22	0.8	2608	1	US-08-160-861-1	Sequence 1, Appli
c	39	22	0.8	3138	1	US-07-867-106-4	Sequence 4, Appli
	40	22	0.8	5852	1	US-07-867-106-2	Sequence 2, Appli
c	41	22	0.8	5852	1	US-07-867-106-2	Sequence 2, Appli
	42	22	0.8	19307	3	US-08-836-022A-10	Sequence 10, Appl
	43	22	0.8	19307	3	US-09-427-048A-10	Sequence 10, Appl
c	44	22	0.8	26664	4	US-09-564-805-28	Sequence 28, Appl
c	45	22	0.8	55827	4	US-09-813-133A-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-07-847-010-18/c
; Sequence 18, Application US/07847010
; Patent No. 5693495
; GENERAL INFORMATION:
;    APPLICANT:  Breiteneder, Heimo
;    APPLICANT:  Reikerstorfer, Arnold
;    APPLICANT:  Valenta, Rudolf
;    APPLICANT:  Hoffmann - Sommergruber, Karin
;    APPLICANT:  Breitenbach, Michael
;    APPLICANT:  Kraft, Dietrich
;    APPLICANT:  Rumpold, Helmut
;    APPLICANT:  Scheiner, Otto
;    APPLICANT:  Ebner, Christof
;    APPLICANT:  Ferreira, Fatima
;    TITLE OF INVENTION:  Allergens of Alder Pollen and
;    TITLE OF INVENTION:  Applications Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Pennie & Edmonds
;       STREET:  1155 Avenue of the Americas
;       CITY:  New York
;       STATE:  New York
;       COUNTRY:  U.S.A.
;       ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/847,010
;       FILING DATE:  01-JUN-1992
;       CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;       NAME:  Jones III, Harry C
;       REGISTRATION NUMBER:  20,280
;       REFERENCE/DOCKET NUMBER:  6530-010
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (212) 790-9090
;       TELEFAX:  (212) 869-9741/8864
;       TELEX:  66141 PENNIE
; INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  860 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
```

;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;   ORGANISM: hazel (Corylus sp.)
;   IMMEDIATE SOURCE:
;   LIBRARY: POLLEN FROM ALLERGON AB, ENGELHOLM, SWEDEN
US-07-847-010-18

Query Match   1.0%;   Score 26;   DB 1;   Length 860;
Best Local Similarity   100.0%;   Pred. No. 0.17;
Matches   26;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   2654 AAATTGACAAAAAAAAAAAAAAAAA 2679
                  |||||||||||||||||||||||||
Db   754 AAATTGACAAAAAAAAAAAAAAAAA 729

RESULT 2
US-09-522-714-15
;   Sequence 15, Application US/09522714
;   Patent No. 6563020
;   GENERAL INFORMATION:
;   APPLICANT: Simmons, Carl R.
;   APPLICANT: Yalpani, Nasser
;   TITLE OF INVENTION: Maize Chitinases and Their Use in
;   TITLE OF INVENTION: Enhancing Disease Resistance in Crop Plants
;   FILE REFERENCE: 1100
;   CURRENT APPLICATION NUMBER: US/09/522,714
;   CURRENT FILING DATE: 2000-03-10
;   EARLIER APPLICATION NUMBER: 60/125,915
;   EARLIER FILING DATE: 1999-03-24
;   NUMBER OF SEQ ID NOS: 32
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 15
;   LENGTH: 1085
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (46)...(966)
US-09-522-714-15

Query Match   0.9%;   Score 25;   DB 4;   Length 1085;
Best Local Similarity   100.0%;   Pred. No. 0.43;
Matches   25;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   2655 AATTGACAAAAAAAAAAAAAAAAAAAA 2679
                  ||||||||||||||||||||||||||
Db   1061 AATTGACAAAAAAAAAAAAAAAAAAAA 1085

RESULT 3
5198347-3
;Patent No. 5198347
;   APPLICANT: Miller, LOUIS H.;ADAMS, JOHN H.;KASLOW,
;DAVIC C.;FANG, XIANGDOUG
;   TITLE OF INVENTION: DNA ENCODING PLASMODIUM VIVAX AND
;PLASMODIUM KNOWLESI DUFFY RECEPTOR
;   NUMBER OF SEQUENCES: 27
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/554,837
;   FILING DATE: 20-JUL-1990
;SEQ ID NO:3:
;   LENGTH: 3157
5198347-3

Query Match   0.9%;   Score 24;   DB 6;   Length 3157;
Best Local Similarity   100.0%;   Pred. No. 1.1;
Matches   24;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   2656 ATTGACAAAAAAAAAAAAAAAAAAAAA 2679
                  |||||||||||||||||||||||||
Db   3088 ATTGACAAAAAAAAAAAAAAAAAAAAA 3111

RESULT 4
US-09-621-976-17572
;   Sequence 17572, Application US/09621976
;   Patent No. 6639063
;   GENERAL INFORMATION:
;   APPLICANT: Dumas Milne Edwards, J.B.
;   APPLICANT: Jobert, S.
;   APPLICANT: Giordano, J.Y.
;   TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;   FILE REFERENCE: GENSET.054PR2
;   CURRENT APPLICATION NUMBER: US/09/621,976
;   CURRENT FILING DATE: 2000-07-21
;   NUMBER OF SEQ ID NOS: 19335
;   SOFTWARE: Patent.pm
;   SEQ ID NO 17572
;   LENGTH: 443
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-17572

Query Match   0.9%;   Score 23;   DB 4;   Length 443;
Best Local Similarity   100.0%;   Pred. No. 3.1;
Matches   23;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   2657 TTGACAAAAAAAAAAAAAAAAAAAA 2679
                  ||||||||||||||||||||||||
Db   420 TTGACAAAAAAAAAAAAAAAAAAAA 442

RESULT 5
US-09-614-912-161
;   Sequence 161, Application US/09614912
;   Patent No. 6677502
;   GENERAL INFORMATION:
;   APPLICANT: Allen, Steve
;   APPLICANT: Rafalski, Antoni
;   APPLICANT: Orozco, Buddy
;   APPLICANT: Miao, Gou-Hau
;   APPLICANT: Famodu, Omolayo O.
;   APPLICANT: Lee, Jian Ming
;   APPLICANT: Sakai, Hajime
;   APPLICANT: Weng, Zude
;   APPLICANT: Caimi, Perry G
;   APPLICANT: Anderson, Shawn
;   TITLE OF INVENTION: Plant Metabolism Genes
;   FILE REFERENCE: BB1378 US NA
;   CURRENT APPLICATION NUMBER: US/09/614,912
;   CURRENT FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: 60/143,401
;   PRIOR FILING DATE: 1999-07-12
;   PRIOR APPLICATION NUMBER: 60/143,412
;   PRIOR FILING DATE: 1999-07-12
;   PRIOR APPLICATION NUMBER: 60/146,650
;   PRIOR FILING DATE: 1999-07-30
;   PRIOR APPLICATION NUMBER: 60/170,906
;   PRIOR FILING DATE: 1999-12-15
;   PRIOR APPLICATION NUMBER: 60/172,959
;   PRIOR FILING DATE: 1999-12-21
;   PRIOR APPLICATION NUMBER: 60/172,946
;   PRIOR FILING DATE: 1999-12-21
;   NUMBER OF SEQ ID NOS: 204
;   SOFTWARE: Microsoft Office 97
;   SEQ ID NO 161
;   LENGTH: 1185
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-614-912-161

Query Match   0.9%;   Score 23;   DB 4;   Length 1185;
Best Local Similarity   100.0%;   Pred. No. 3;
Matches   23;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy 2657 TTGACAAAAAAAAAAAAAAAAAAA 2679
         ||||||||||||||||||||||||
Db 1163 TTGACAAAAAAAAAAAAAAAAA 1185

RESULT 6
US-09-412-102-1
; Sequence 1, Application US/09412102
; Patent No. 6228992
; GENERAL INFORMATION:
; APPLICANT: JESSEN, HOLLY J
; APPLICANT: MEYER, TERRY E
; TITLE OF INVENTION: GENES AND METHODS FOR CONTROL OF
; TITLE OF INVENTION: NEMATODES IN PLANTS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)
; STREET: 3605 Glenwood Ave. Suite 310
; CITY: Raliegh
; STATE: NC
; COUNTRY: US
; ZIP: 27622
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/412,102
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/217,787
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 5718-18P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919 420 2202
; TELEFAX: 919 881 3175
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1347 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Zea mays
; IMMEDIATE SOURCE:
; CLONE: P12217
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 146..994
; US-09-412-102-1

Query Match   0.9%; Score 23; DB 3; Length 1347;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches   23; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 2657 TTGACAAAAAAAAAAAAAAAAAAA 2679
         ||||||||||||||||||||||||
Db 1325 TTGACAAAAAAAAAAAAAAAAA 1347

RESULT 7
US-09-217-787-1
; Sequence 1, Application US/09217787
; Patent No. 6284948
; GENERAL INFORMATION:
; APPLICANT: JESSEN, HOLLY J
; APPLICANT: MEYER, TERRY E

; TITLE OF INVENTION: GENES AND METHODS FOR CONTROL OF
; TITLE OF INVENTION: NEMATODES IN PLANTS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: W. Murray Spruill (Alston & Bird, LLP)
; STREET: 3605 Glenwood Ave. Suite 310
; CITY: Raliegh
; STATE: NC
; COUNTRY: US
; ZIP: 27622
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/217,787
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; REFERENCE/DOCKET NUMBER: 5718-18P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919 420 2202
; TELEFAX: 919 881 3175
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1347 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Zea mays
; IMMEDIATE SOURCE:
; CLONE: P12217
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 146..994
; US-09-217-787-1

Query Match   0.9%; Score 23; DB 3; Length 1347;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches   23; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy 2657 TTGACAAAAAAAAAAAAAAAAAAA 2679
         ||||||||||||||||||||||||
Db 1325 TTGACAAAAAAAAAAAAAAAAA 1347

RESULT 8
US-08-484-101B-35
; Sequence 35, Application US/08484101B
; Patent No. 5824868
; GENERAL INFORMATION:
; APPLICANT: California Institute of Technology
; TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO
; TITLE OF INVENTION: ETHYLENE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard F. Trecartin
; STREET: 3400 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```
;       APPLICATION NUMBER:  US/08/484,101B
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/
;       FILING DATE:  01-JUL-1994
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/086,555
;       FILING DATE:  01-JUL-1993
;       CLASSIFICATION:  800
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Trecartin, Richard F.
;     REGISTRATION NUMBER:  31,801
;     REFERENCE/DOCKET NUMBER:  A-57515-2/RFT
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (415) 781-1989
;     TELEFAX:  (415) 398-3249
; INFORMATION FOR SEQ ID NO: 35:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  2405 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  288..2196
US-08-484-101B-35

Query Match             0.9%;  Score 23;  DB 1;  Length 2405;
Best Local Similarity   100.0%;  Pred. No. 2.8;
Matches   23;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    2657 TTGACAAAAAAAAAAAAAAAAAAA 2679
           |||||||||||||||||||||||
Db    2378 TTGACAAAAAAAAAAAAAAAAAAA 2400

RESULT 9
US-08-484-101B-49
; Sequence 49, Application US/08484101B
; Patent No. 5824868
; GENERAL INFORMATION:
;   APPLICANT:  California Institute of Technology
;   TITLE OF INVENTION:  PLANTS HAVING MODIFIED RESPONSE TO
;   TITLE OF INVENTION:  ETHYLENE
;   NUMBER OF SEQUENCES:  50
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Richard F. Trecartin
;     STREET:  3400 Embarcadero Center, Suite 3400
;     CITY:  San Francisco
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  94111
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/484,101B
;     FILING DATE:  07-JUN-1995
;     CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/
;       FILING DATE:  01-JUL-1994
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/086,555
;       FILING DATE:  01-JUL-1993
;       CLASSIFICATION:  800
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Trecartin, Richard F.
```

```
;     REGISTRATION NUMBER:  31,801
;     REFERENCE/DOCKET NUMBER:  A-57515-2/RFT
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (415) 781-1989
;     TELEFAX:  (415) 398-3249
; INFORMATION FOR SEQ ID NO: 49:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  2405 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  288..2196
US-08-484-101B-49

Query Match             0.9%;  Score 23;  DB 1;  Length 2405;
Best Local Similarity   100.0%;  Pred. No. 2.8;
Matches   23;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    2657 TTGACAAAAAAAAAAAAAAAAAAA 2679
           |||||||||||||||||||||||
Db    2378 TTGACAAAAAAAAAAAAAAAAAAA 2400

RESULT 10
US-08-714-524D-35
; Sequence 35, Application US/08714524D
; Patent No. 6294716
; GENERAL INFORMATION:
;   APPLICANT:  Meyerowitz, Elliott M
;   APPLICANT:  Chang, Caren
;   APPLICANT:  Bleecker, Anthony B
;   TITLE OF INVENTION:  PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
;   FILE REFERENCE:  a-57515-4
;   CURRENT APPLICATION NUMBER:  US/08/714,524D
;   CURRENT FILING DATE:  1996-09-16
;   NUMBER OF SEQ ID NOS:  56
;   SOFTWARE:  PatentIn Ver. 2.1
SEQ ID NO 35
;   LENGTH:  2405
;   TYPE:  DNA
;   ORGANISM:  Lycopersicon esculentum
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  (288)..(2195)
US-08-714-524D-35

Query Match             0.9%;  Score 23;  DB 3;  Length 2405;
Best Local Similarity   100.0%;  Pred. No. 2.8;
Matches   23;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    2657 TTGACAAAAAAAAAAAAAAAAAAA 2679
           |||||||||||||||||||||||
Db    2378 TTGACAAAAAAAAAAAAAAAAAAA 2400

RESULT 11
US-08-714-524D-49
; Sequence 49, Application US/08714524D
; Patent No. 6294716
; GENERAL INFORMATION:
;   APPLICANT:  Meyerowitz, Elliott M
;   APPLICANT:  Chang, Caren
;   APPLICANT:  Bleecker, Anthony B
;   TITLE OF INVENTION:  PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
;   FILE REFERENCE:  a-57515-4
;   CURRENT APPLICATION NUMBER:  US/08/714,524D
;   CURRENT FILING DATE:  1996-09-16
;   NUMBER OF SEQ ID NOS:  56
;   SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 49
;   LENGTH:  2405
```

; TYPE: DNA
; ORGANISM: Lycopersicon esculentum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (288)..(2195)
US-08-714-524D-49

Query Match   0.9%; Score 23; DB 3; Length 2405;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches   23; Conservative 0; Mismatches   0; Indels   0; Gaps   0;

Qy   2657 TTGACAAAAAAAAAAAAAAAAAAAA 2679
          ||||||||||||||||||||||||
Db   2378 TTGACAAAAAAAAAAAAAAAAAAAA 2400

RESULT 12
US-08-244-113-18
; Sequence 18, Application US/08244113
; Patent No. 5455181
; GENERAL INFORMATION:
; APPLICANT: Strube, Karl-Hermann
; APPLICANT: Bialojan, Siegfried
; APPLICANT: Kroeger, Burkhard
; APPLICANT: Friedrich, Thomas
; TITLE OF INVENTION: No. 5455181el thrombin-inhibitory proteins from terrestrial
; TITLE OF INVENTION: leeches.
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,113
; FILING DATE:
; CLASSIFICATION: 530
; CLASSIFICATION: C07K 73/10
; CLASSIFICATION: A61K 37/64
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP92/02661
; FILING DATE: 19-NOV-1992
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-244-113-18

Query Match   0.8%; Score 22; DB 1; Length 277;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches   22; Conservative 0; Mismatches   0; Indels   0; Gaps   0;

Qy   2658 TGACAAAAAAAAAAAAAAAAAAAA 2679
          |||||||||||||||||||||||
Db   224 TGACAAAAAAAAAAAAAAAAAAAA 245

RESULT 13
US-09-280-116-159/c
; Sequence 159, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs

; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 159
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Lon family of ATP-dependent proteases
US-09-280-116-159

Query Match   0.8%; Score 22; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches   22; Conservative 0; Mismatches   0; Indels   0; Gaps   0;

Qy   2658 TGACAAAAAAAAAAAAAAAAAAAA 2679
          |||||||||||||||||||||||
Db   22 TGACAAAAAAAAAAAAAAAAAAAA 1

RESULT 14
US-09-205-258-172
; Sequence 172, Application US/0920525 8
; Patent No. 6525174
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: PZ007P1
; CURRENT APPLICATION NUMBER: US/09/205,258
; CURRENT FILING DATE: 1999-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,885
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,880
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,971
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,882
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,899
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,893
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,900
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,901
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,892
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,915
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,019

```
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,972
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,916
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,373
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,875
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,374
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,917
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,949
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,883
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,897
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,898
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,962
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,963
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,877
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,878
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/070,923
EARLIER FILING DATE: 1997-12-18
EARLIER APPLICATION NUMBER: 60/092,921
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/094,657
EARLIER FILING DATE: 1998-07-30
NUMBER OF SEQ ID NOS: 1227
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 172
LENGTH: 786
TYPE: DNA
ORGANISM: Homo sapiens
US-09-205-258-172

Query Match          0.8%;  Score 22;  DB 4;  Length 786;
Best Local Similarity 100.0%;  Pred. No. 7.9;
Matches   22;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy  2658 TGACAAAAAAAAAAAAAAAAAA 2679
         ||||||||||||||||||||||
Db   760 TGACAAAAAAAAAAAAAAAAAA 781


RESULT 15
US-09-205-258-190/c
Sequence 190, Application US/09205258
Patent No. 6525174
GENERAL INFORMATION:
APPLICANT: Young et al.
TITLE OF INVENTION: 207 Human Secreted Proteins
FILE REFERENCE: PZ007P1
CURRENT APPLICATION NUMBER: US/09/205,258
CURRENT FILING DATE: 1998-12-04
CURRENT APPLICATION NUMBER: PCT/US98/11422
EARLIER FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 60/048,885
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,375
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,881
```

```
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,880
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,896
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,020
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,876
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,895
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,884
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,894
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,971
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,964
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,882
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,899
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,893
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,900
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,901
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,892
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,915
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,019
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,972
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,916
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,373
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,875
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/049,374
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,917
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,883
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,877
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/048,878
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/070,923
EARLIER FILING DATE: 1997-12-18
EARLIER APPLICATION NUMBER: 60/092,921
EARLIER FILING DATE: 1998-07-15
EARLIER APPLICATION NUMBER: 60/094,657
EARLIER FILING DATE: 1998-07-30
```

```
;   NUMBER OF SEQ ID NOS: 1227
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 190
;     LENGTH: 906
;     TYPE: DNA
;     ORGANISM: Homo sapiens
;     FEATURE:
;     NAME/KEY: SITE
;     LOCATION: (144)
;     OTHER INFORMATION: n equals a,t,g, or c
;     FEATURE:
;     NAME/KEY: SITE
;     LOCATION: (145)
;     OTHER INFORMATION: n equals a,t,g, or c
US-09-205-258-190

Query Match         0.8%;  Score 22;  DB 4;  Length 906;
Best Local Similarity   100.0%;  Pred. No. 7.9;
Matches   22;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    2657 TTGACAAAAAAAAAAAAAAAA 2678
           ||||||||||||||||||||
Db     843 TTGACAAAAAAAAAAAAAAAA 822

Search completed: May 15, 2004, 08:02:31
Job time : 216 secs
```

```
          GenCore version 5.1.6
   Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May 7, 2004, 14:49:13 ; Search time 11.5528 Seconds
                           (without alignments)
                           1752.371 Million cell updates/sec

Title:     US-10-048-046-2_COPY_31_103
Perfect score: 378
Sequence: 1 VLLRKREWTIGRRRGCDLSF........INKLKVVKKQTCPLQTGDVI 73

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:  1140673 seqs, 27756755 residues

Total number of hits satisfying chosen parameters:    1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Published_Applications_AA:*
   1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
   2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
   3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
   4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
   5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
   6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
   7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
   8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
   9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
  10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
  11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
  12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
  13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
  14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
  15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
  16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
  17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
  18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	378	100.0	426	9	US-09-764-864-809	Sequence 809, App
2	378	100.0	664	9	US-09-780-525-2	Sequence 2, Appli
3	304	80.4	92	9	US-09-764-864-1109	Sequence 1109, Ap
4	275	72.8	92	9	US-09-764-864-1110	Sequence 1110, Ap
5	275	72.8	99	9	US-09-764-864-1529	Sequence 1529, Ap
6	107.5	28.4	522	9	US-09-740-627-11	Sequence 11, Appl
7	107.5	28.4	821	14	US-10-081-119-16	Sequence 16, Appl
8	99	26.2	699	14	US-10-032-585-7635	Sequence 7635, Ap
9	81	21.4	445	12	US-10-424-599-178873	Sequence 178873,
10	81	21.4	1082	12	US-10-425-114-56100	Sequence 56100, A
11	78.5	20.8	175	9	US-09-867-550-1026	Sequence 1026, Ap
12	78.5	20.8	244	9	US-09-764-864-1111	Sequence 1111, Ap
13	77.5	20.5	460	9	US-09-740-627-10	Sequence 10, Appli
14	76.5	20.2	517	15	US-10-618-173-2	Sequence 2, Appli
15	74.5	19.7	229	12	US-10-424-599-187236	Sequence 187236,
16	74.5	19.7	376	12	US-10-424-599-233512	Sequence 233512,
17	74	19.6	250	12	US-10-424-599-269525	Sequence 269525,
18	72	19.0	513	15	US-10-369-493-1578	Sequence 1578, Ap
19	71	18.8	409	12	US-10-282-122A-64192	Sequence 64192, A
20	70.5	18.7	874	12	US-10-282-122A-45185	Sequence 45185, A
21	67	17.7	481	14	US-10-156-761-12906	Sequence 12906, A
22	66.5	17.6	288	9	US-09-738-626-3552	Sequence 3552, Ap
23	66	17.5	435	12	US-10-424-599-276420	Sequence 276420,
24	64	16.9	545	14	US-10-142-356-6	Sequence 6, Appli
25	63.5	16.8	474	15	US-10-369-493-2299	Sequence 2299, Ap
26	63.5	16.8	492	15	US-10-369-493-3607	Sequence 3607, Ap
27	63	16.7	138	14	US-10-238-075-1035	Sequence 1035, Ap
28	62.5	16.5	527	9	US-09-712-363-156	Sequence 156, App
29	62.5	16.5	527	14	US-10-080-170-348	Sequence 348, App
30	62	16.4	467	12	US-10-406-031-5	Sequence 5, Appli
31	62	16.4	1724	12	US-10-276-774-2176	Sequence 2176, Ap
32	61.5	16.3	207	9	US-09-740-627-9	Sequence 9, Appli
33	61.5	16.3	459	9	US-10-282-122A-49959	Sequence 49959, A
34	61.5	16.3	488	14	US-10-080-170-4	Sequence 4, Appli
35	61	16.1	1055	12	US-10-282-122A-57704	Sequence 57704, A
36	61	16.1	1056	12	US-10-282-122A-57499	Sequence 57499, A
37	60.5	16.0	365	9	US-09-738-626-4872	Sequence 4872, Ap
38	60.5	16.0	397	12	US-10-389-647-430	Sequence 430, App
39	60	15.9	939	12	US-10-282-122A-47696	Sequence 47696, A
40	60	15.9	969	12	US-10-282-122A-63852	Sequence 63852, A
41	59.5	15.7	804	12	US-10-282-122A-42499	Sequence 42499, A
42	59.5	15.7	820	9	US-09-815-242-10771	Sequence 10771, A
43	59	15.6	601	12	US-10-267-502-367	Sequence 367, App
44	59	15.6	1045	12	US-10-041-018-199	Sequence 199, App
45	59	15.6	1045	12	US-10-041-018-218	Sequence 218, App

ALIGNMENTS

```
RESULT 1
US-09-764-864-809
; Sequence 809, Application US/09764864
; Patent No. US2002013275A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PT223
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE: 2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 809
;  LENGTH: 426
;  TYPE: PRT
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: SITE
;  LOCATION: (414)
;  OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;  NAME/KEY: SITE
;  LOCATION: (415)
;  OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;  NAME/KEY: SITE
;  LOCATION: (420)
;  OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-809

Query Match           100.0%;   Score 378;  DB 9;  Length 426;
Best Local Similarity 100.0%;   Pred. No. 6.9e-39;
Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 92

Qy   61 KKQTCPLQTGDVI 73
```

(handwritten annotations: "not 102?", "(title of the 2000", "80/2")

Db   93 KKQTCPLQTGDVI 105
         |||||||||||||

RESULT 2
US-09-780-525-2
; Sequence 2, Application US/09780525
; Patent No. US2002004223A1
; GENERAL INFORMATION:
; APPLICANT: Bin-Bing Zhou
; APPLICANT: Yuan Zhu
; APPLICANT: Priya Chaturvedi
; APPLICANT: Mark R. Hurle
; APPLICANT: Xiaotong Li
; TITLE OF INVENTION: FHAR1, A NEW RING FINGER PROTEIN
; FILE REFERENCE: GP-70668-C1
; CURRENT APPLICATION NUMBER: US/09/780,525
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 09/456,876
; PRIOR FILING DATE: 1999-12-08
; SOFTWARE: FastSEQ for Windows Version 3.0
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 664
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-780-525-2

Query Match   100.0%; Score 378; DB 9; Length 664;
Best Local Similarity 100.0%; Pred. No. 1.2e-38;
Matches   73; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 90

Qy   61 KKQTCPLQTGDVI 73
         |||||||||||||
Db   91 KKQTCPLQTGDVI 103

RESULT 3
US-09-764-864-1109
; Sequence 1109, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1109
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1109

Query Match   80.4%; Score 304; DB 9; Length 92;
Best Local Similarity   98.3%; Pred. No. 2.1e-30;
Matches   59; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTXTSGTVINKLKVV 92

RESULT 4
US-09-764-864-1110
; Sequence 1110, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1110
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (17)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (96)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1110

Query Match   72.8%; Score 275; DB 9; Length 99;
Best Local Similarity   90.3%; Pred. No. 1e-26;
Matches   56; Conservative   1; Mismatches   3; Indels   2; Gaps   1;

Qy   12 RRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 71
         | :|| || |||||||||||||||||||||||||||||||||||||||||||
Db   9 RDKGPD--FPXNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 66

Qy   72 VI 73
         ||
Db   67 VI 68

RESULT 5
US-09-764-864-1529
; Sequence 1529, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1529
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (17)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (96)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1529

Query Match   72.8%; Score 275; DB 9; Length 99;
Best Local Similarity   90.3%; Pred. No. 1e-26;
Matches   56; Conservative   1; Mismatches   3; Indels   2; Gaps   1;

Qy   12 RRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 71
         | :|| || ||||||||||||||||||||||||||||||||||||||||||
Db   9 RDKGPD--FPXNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 66

Qy   72 VI 73
         | |
Db   67 VI 68

RESULT 6
US-09-740-627-11
; Sequence 11, Application US/09740627
; Patent No. US2002012964A1
; GENERAL INFORMATION:
;   APPLICANT: Nakanishi, Makoto
;   TITLE OF INVENTION: CELL CYCLE REGULATING FACTOR
;   FILE REFERENCE: 06501-071001
;   CURRENT APPLICATION NUMBER: US/09/740,627
;   CURRENT FILING DATE: 2000-12-19
;   PRIOR APPLICATION NUMBER: PCT/JP99/03350
;   PRIOR FILING DATE: 1999-06-23
;   PRIOR APPLICATION NUMBER: JP 10/192467
;   PRIOR FILING DATE: 1998-06-23
;   NUMBER OF SEQ ID NOS: 11
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
;   LENGTH: 522
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-740-627-11

Query Match   28.4%;   Score 107.5; DB 9; Length 522;
Best Local Similarity   37.7%;   Pred. No. 0.0001;
Matches   26; Conservative   9; Mismatches 33; Indels   1; Gaps   1;

Qy   5 KREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQT 64
           |:||:|||| | : :|:|||||| :|| |:||||:||:|| |:| :: |
Db   63 KKVWTFGRNPACDYHLGNISRLSNKHFQILLGE-DGNLLLNDISTNGTWLNGQKVEKNSN 121

Qy   65 CPLQTGDVI 73
           | | |
Db   122 QLLSQGDEI 130

RESULT 7
US-10-081-119-16
; Sequence 16, Application US/10081119
; Publication No. US2003004549lA1
; GENERAL INFORMATION:
;   APPLICANT: Reinhard, Christoph
;   APPLICANT: Jefferson, Anne B.
;   APPLICANT: Chan, Vivien W.
;   TITLE OF INVENTION: TTK in Diagnosis and as a Therapeutic
;   TITLE OF INVENTION: Target in Cancer
;   FILE REFERENCE: 16932.002
;   CURRENT APPLICATION NUMBER: US/10/081,119
;   CURRENT FILING DATE: 2002-02-21
;   PRIOR APPLICATION NUMBER: 60/289,813
;   PRIOR FILING DATE: 2001-02-21
;   NUMBER OF SEQ ID NOS: 38
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 821
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-10-081-119-16

Query Match   28.4%;   Score 107.5; DB 14; Length 821;
Best Local Similarity   37.7%;   Pred. No. 0.00018;
Matches   26; Conservative   9; Mismatches 33; Indels   1; Gaps   1;

Qy   5 KREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQT 64
           |:||:|||| | : :|:|||||| :|| |:||||:||:|| |:| :: |
Db   63 KKVWTFGRNPACDYHLGNISRLSNKHFQILLGE-DGNLLLNDISTNGTWLNGQKVEKNSN 121

Qy   65 CPLQTGDVI 73
           | | |

Db   122 QLLSQGDEI 130

RESULT 8
US-10-032-585-7635
; Sequence 7635, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
;   APPLICANT: Terry, Roemer D.
;   APPLICANT: Bo, Jiang
;   APPLICANT: Charles, Boone
;   APPLICANT: Howard, Bussey
;   TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
;   FILE REFERENCE: 10182-005-999
;   CURRENT APPLICATION NUMBER: US/10/032,585
;   CURRENT FILING DATE: 2001-12-20
;   NUMBER OF SEQ ID NOS: 8000
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7635
;   LENGTH: 699
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-10-032-585-7635

Query Match   26.2%;   Score 99; DB 14; Length 699;
Best Local Similarity   33.3%;   Pred. No. 0.0017;
Matches   23; Conservative 12; Mismatches 34; Indels   0; Gaps   0;

Qy   5 KREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQT 64
           : |||| :| | : ||:| :::|:| :|| |:||||:||:|| :: ::: ::||
Db   59 KMTWYFGRDPNSDLQVASSRISNKHFQIWLNFNDKSLWIKDTSTNGTHLNNSRLVKGSN 118

Qy   65 CPLQTGDVI 73
           || | |
Db   119 YLLNQGDEI 127

RESULT 9
US-10-424-599-178873
; Sequence 178873, Application US/10424599
; Publication No. US2004003l072Al
; GENERAL INFORMATION:
;   APPLICANT: La Rosa Thomas J
;   APPLICANT: Kovalic David K
;   APPLICANT: Zhou Yihua
;   APPLICANT: Cao Yongwei
;   TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53223)B
;   CURRENT APPLICATION NUMBER: US/10/424,599
;   CURRENT FILING DATE: 2003-04-28
;   NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 178873
;   LENGTH: 445
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_132539C.1.pep
US-10-424-599-178873

Query Match   21.4%;   Score 81; DB 12; Length 445;
Best Local Similarity   29.6%;   Pred. No. 0.18;
Matches   24; Conservative 10; Mismatches 27; Indels 20; Gaps   2;

Qy   13 RRGCDLSFPSNKLVSGDHCRIVVD----------EKSGQVTLEDTSTSGTVINKLKV 59
           :|:||: || |||| |||:|| : :: :|:|:::|::|| ::
Db   29 RKGGDVIITKDKGVSRVHAEIVNTVNPVNPLPNERSHLSSSIHIRDCSKYGTFINKNGG 88

Qy   60 VKKQT------CPLQTGDVI 73
           : || ||| | | |
Db   89 AKKKVHELPNKETALENGDLV 109

RESULT 10
US-10-425-114-56100
; Sequence 56100, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
US-10-425-114-56100
; SEQ ID NO 56100
; LENGTH: 1082
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: JC-GMLE01810018D04_FLI.pep
US-10-425-114-56100

Query Match   21.4%;   Score 81; DB 12; Length 1082;
Best Local Similarity   33.8%;   Pred. No. 0.55;
Matches   23; Conservative   12; Mismatches   27; Indels   6; Gaps   4;

Qy   10 IGRRRGCDLSFPSN-KLVSGDHCRIVDEKSGGQ--VTLEDTSTSGTV-INKLKVVKKQTC 65
              |||| : |: |::|: ||: :: |:|| :| : :||| ||:|||: :|
Db   1 IGSNRSC--NFPLNDQTISGNLCKIKHTQGDGSAVAVLESMGSKGSVLVNGTHVAKNTSC 58

Qy   66 PLQTGDVI 73
              :|||
Db   59 VLNSGDEV 66

RESULT 11
US-09-867-550-1026
; Sequence 1026, Application US/09867550
; Patent No. US2002008220A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US2002008220A1e1 Polynucleotides from Atherogenic Cells and
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1026
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-867-550-1026

Query Match   20.8%;   Score 78.5; DB 9; Length 175;
Best Local Similarity   32.7%;   Pred. No. 0.12;
Matches   18; Conservative   15; Mismatches   17; Indels   5; Gaps   2;

Qy   24 KLVSGDHCRIVDEKSGQVTLEDT-STSGTVINKLKVVK----KQTCPLQTGDVI 73
              |::|: :| |:|::| :||||| ||||||||| ::| : |:||
Db   49 KVLSRNHALVWFDHKTGKFYLQDTKSSNGTFINSQRLSRGSEESPPCEILSGDII 103

RESULT 12

US-09-764-864-1111
; Sequence 1111, Application US/09764864
; Patent No. US2002013275A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1111
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (74)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (159)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (203)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (229)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1111

Query Match   20.8%;   Score 78.5; DB 9; Length 244;
Best Local Similarity   32.7%;   Pred. No. 0.18;
Matches   18; Conservative   15; Mismatches   17; Indels   5; Gaps   2;

Qy   24 KLVSGDHCRIVDEKSGQVTLEDT-STSGTVINKLKVVK----KQTCPLQTGDVI 73
              |::|: :| |:|::| :||||| ||||||||| ::| : |:||
Db   77 KVLSRNHALVWFDHKTGKFYLQDTKSSNGTFINSQRLSRGSEESPPCEILSGDII 131

RESULT 13
US-09-740-627-10
; Sequence 10, Application US/09740627
; Patent No. US2002001296A1
; GENERAL INFORMATION:
; APPLICANT: Nakanishi, Makoto
; TITLE OF INVENTION: CELL CYCLE REGULATING FACTOR
; FILE REFERENCE: 06501-071001
; CURRENT APPLICATION NUMBER: US/09/740,627
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: PCT/JP99/03350
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: JP 10/192467
; PRIOR FILING DATE: 1998-06-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-09-740-627-10

Query Match   20.5%;   Score 77.5; DB 9; Length 460;
Best Local Similarity   30.6%;   Pred. No. 0.52;
Matches   22; Conservative   9; Mismatches   34; Indels   7; Gaps   2;

Qy   8 WTIGRRRGCDLSFPSNKLVSGDHCRIV------VDEKSGGQVTLEDTSTSGTVINKLKVVK 61
              | |||: :| |: :|| |:| : | |:|: |: | ::: || |:|||
Db   60 WRFGRHKSCEVVL-NGPRVSNFHFEIYQGHRNDSDESENVVFLHDHSSNGTFLNFERLAK 118

Qy   62 KQTCPLQTGDVI 73
              | | | |
Db   119 NSRTILSNGDEI 130

```
RESULT 14
US-10-618-173-2
; Sequence 2, Application US/10618173
; Publication No. US20040018603A1
; GENERAL INFORMATION:
;   APPLICANT: CALIFORNIA INSTITUTE OF TECHNOLOGY
;   APPLICANT: GUO, Zijian
;   APPLICANT: DUNPHY, William
;   TITLE OF INVENTION: CHECKPOINT-ACTIVATING OLIGONUCLEOTIDES
;   FILE REFERENCE: CIT1350-1
;   CURRENT APPLICATION NUMBER: US/10/618,173
;   CURRENT FILING DATE: 2003-07-11
;   PRIOR APPLICATION NUMBER: US/09/849,617
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US 60/202,028
;   PRIOR FILING DATE: 2000-05-04
;   NUMBER OF SEQ ID NOS: 14
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 517
;   TYPE: PRT
;   ORGANISM: Xenopus laevis
US-10-618-173-2

Query Match            20.2%;  Score 76.5;  DB 15;  Length 517;
Best Local Similarity  31.9%;  Pred. No. 0.81;
Matches    23;  Conservative   9;  Mismatches  29;  Indels  11;  Gaps   2;

Qy        7  EWTIGRRRGCDLSF-------PSNKLVSGDHCRIVVDEKSGQ--VTLEDTSTSGTVIN 55
              |:||:||| :|            ::|||| |::|||| :|| :|
Db       84  EYVFGRDKKCDYTFDIPVLNQTDRYKTYSKRHFRIFQELGHGHSRVANIEDLSGNGTFVN 143

Qy       56  KLKVVKKQTCPL 67
              |:| :||
Db      144  KEIIGKGRTLPL 155


RESULT 15
US-10-424-599-187236
; Sequence 187236, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa Thomas J
;   APPLICANT: Kovalic David K
;   APPLICANT: Zhou Yihua
;   APPLICANT: Cao Yongwei
;   TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53223)B
;   CURRENT APPLICATION NUMBER: US/10/424,599
;   CURRENT FILING DATE: 2003-04-28
;   NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 187236
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(229)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_140086C.1.pep
US-10-424-599-187236

Query Match            19.7%;  Score 74.5;  DB 12;  Length 229;
Best Local Similarity  36.0%;  Pred. No. 0.52;
Matches    27;  Conservative   9;  Mismatches  32;  Indels   7;  Gaps   4;

Qy        2  LLRKREWTIGRRRGCD--LSFPSNKLVSGDHCRIVVDEKSGQVTLEDT-STSGTVINKLK 58
              :  :  |  |  |||   :| :|  ::|||||| |::||| |  | ||:|| :
Db       98  VFEKGAYMFGRLDLCDFVLEHPT---ISRFHA-VVQFKRSGDAYLYDLGSTHGTFLNKNQ 153


Qy       59  VVKKQTCPLQTGDVI 73
              |  |   |  |  ||||
Db      154  VEKNTYVDLHVGDVI 168

Search completed: May 7, 2004, 15:06:50
Job time : 12.5628 secs
```

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    May 14, 2004, 17:19:10 ; Search time 1591 Seconds
                                    (without alignments)
                        7641.379 Million cell updates/sec

Title:          US-10-048-046-1
Perfect score:  2679
Sequence:       1 aagaattcggcacgaggccg.........acaaaaaaaaaaaaaaaa 2679

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2947324 seqs, 226904515 residues

Total number of hits satisfying chosen parameters:        5894648

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      Published Applications NA:*
    1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
    2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
    3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
    4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
    5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
    6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
    7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
    8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
    9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
   10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
   11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
   12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
   13:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
   14:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
   15:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
   16:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
   17:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
   18:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
   19:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                        SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1995	74.5	1995	9	US-09-780-525-1	Sequence 1, Appli
2	1919.4	71.6	2448	16	US-10-108-260A-2073	Sequence 2073, Ap
3	945	35.3	1311	9	US-09-764-864-22	Sequence 22, Appl
4	876.2	32.7	2186	16	US-10-094-749-160	Sequence 160, App
5	536.6	20.0	693	9	US-09-764-864-483	Sequence 483, App
6	394.4	14.7	476	10	US-09-918-995-2180	Sequence 2180, Ap
7	310.2	11.6	357	9	US-09-764-864-322	Sequence 322, App
8	273.6	10.2	449	10	US-09-918-995-13662	Sequence 13662, A
9	256.4	9.6	354	9	US-09-764-864-323	Sequence 323, App
10	256.4	9.6	354	9	US-09-764-864-742	Sequence 742, App
11	60	2.2	60	10	US-09-908-975-6220	Sequence 6220, Ap
12	51.4	1.9	1255	13	US-10-424-599-95539	Sequence 95539, A
13	51	1.9	383	10	US-09-814-353-18006	Sequence 18006, A
14	51	1.9	411	13	US-10-424-599-127666	Sequence 127666,

						ID		Description
c	15	49	1.8	434	9	US-09-960-352-3525		Sequence 3525, Ap
c	16	48.2	1.8	7455	15	US-10-311-455-1732		Sequence 1732, Ap
c	17	47.8	1.8	312	9	US-09-960-352-8414		Sequence 8414, Ap
c	18	47.2	1.8	424	9	US-09-960-352-11218		Sequence 11218, A
c	19	47.2	1.8	6215	15	US-10-311-455-1163		Sequence 1163, Ap
c	20	47	1.7	349	10	US-09-814-353-18039		Sequence 18039, A
c	21	46.6	1.7	1290	15	US-10-321-204-8		Sequence 8, Appli
c	22	46.6	1.7	12590	15	US-10-311-455-1893		Sequence 1893, Ap
c	23	46.4	1.7	388	10	US-09-814-353-17525		Sequence 17525, A
c	24	46.4	1.7	425	9	US-09-834-975-451		Sequence 451, App
c	25	46.2	1.7	198	13	US-09-732-560-26		Sequence 26, Appl
c	26	46.2	1.7	198	13	US-09-732-560-26		Sequence 26, Appl
c	27	46.2	1.7	312	9	US-09-960-352-8414		Sequence 8414, Ap
c	28	46.2	1.7	373	9	US-09-960-352-836		Sequence 836, App
c	29	46.2	1.7	19653	15	US-10-311-455-1308		Sequence 1308, Ap
c	30	46	1.7	600	15	US-10-198-846-8434		Sequence 8434, Ap
c	31	46	1.7	17869	15	US-10-311-455-78		Sequence 78, Appl
c	32	46	1.7	17869	17	US-10-257-166-2		Sequence 2, Appli
c	33	45.8	1.7	244	13	US-10-424-599-12849		Sequence 12849, A
c	34	45.8	1.7	613	13	US-10-424-599-34242		Sequence 34242, A
c	35	45.8	1.7	6219	13	US-10-240-454-20		Sequence 20, Appl
c	36	45.8	1.7	6219	15	US-10-311-455-840		Sequence 840, App
c	37	45.4	1.7	358	15	US-10-198-846-9603		Sequence 9603, Ap
c	38	45.4	1.7	440	13	US-10-085-783A-376		Sequence 376, App
c	39	45.4	1.7	440	16	US-10-242-535A-376		Sequence 376, App
c	40	45.2	1.7	384	10	US-09-814-353-18604		Sequence 18604, A
c	41	45.2	1.7	496	10	US-09-814-353-4862		Sequence 4862, Ap
c	42	45.2	1.7	496	10	US-09-814-353-11159		Sequence 11159, A
c	43	45.2	1.7	1803	13	US-10-424-599-40634		Sequence 40634, A
c	44	45.2	1.7	69350	12	US-10-041-018-379		Sequence 379, App
c	45	45	1.7	638	15	US-10-198-846-8560		Sequence 8560, Ap

```
                        ALIGNMENTS

RESULT 1
US-09-780-525-1
; Sequence 1, Application US/09780525
; Patent No. US2002004223A1
; GENERAL INFORMATION:
;  APPLICANT: Bin-Bing Zhou
;  APPLICANT: Yuan Zhu
;  APPLICANT: Priya Chaturvedi
;  APPLICANT: Mark R. Hurle
;  APPLICANT: Xiaotong Li
;  TITLE OF INVENTION: FHAR1, A NEW RING FINGER PROTEIN
;  FILE REFERENCE: GP-70668-C1
;  CURRENT APPLICATION NUMBER: US/09/780,525
;  CURRENT FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 09/456,876
;  PRIOR FILING DATE: 1999-12-08
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;  LENGTH: 1995
;  TYPE: DNA
;  ORGANISM: HOMO SAPIENS
US-09-780-525-1

Query Match             74.5%;  Score 1995;  DB 9;  Length 1995;
Best Local Similarity   74.5%;  Pred. No. 0;
Matches 1995;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    91  ATCGGAGCGGCCGAGGAAGGCAAGCAGTCGCGCCGCGCCGCCCTGGGACCGGCTCCTG  150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  ATCGGAGCGGCCGAGGAAGGCAAGCAGTCGCGCCGCGCCGCCCTGGGACCGGCTCCTG   60

Qy   151  CGTCTGGGCGCGGAGGACGTTCCTCCTGAGGAAGCGGGAGTTGGACCATC  210
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  CGTCTGGGCGCGGAGGACGTTCCTCCTGAGGAAGCGGGAGTTGGACCATC  120

Qy   211  GGGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGATCAC  270
```

RESULT 2
US-10-108-260A-2073
; Sequence 2073, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2073
; LENGTH: 2448
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-2073

Query Match
Best Local Similarity   71.6%;   Score 1919.4;   DB 16;   Length 2448;
Matches 2227;   Conservative 88.5%;   Pred. No. 0;   Mismatches 6;   Indels 284;   Gaps 2;

```
Qy    218   GACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTCGTCTCTGGAGATCACTGTAGAA    277
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    887   ATAAAATCACTCAAGAGACCTGCTGAGAGCTGTCAGCGCCCAAAGTGCAGTCAGGCGGCGTCTTTTTCTGATGAAGAAG    946

Db     83   GACCCTGTTGCGGCACCTTTCCTTCCCCAGCAATAAACTCGTCTCTGGAGATCACTGTAGAA    142
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1298   GGAGTTCAGGAGACCTGCTGAGAGCTGTCAGTGACAGTGAGTCCTCAGACATTAGCC    1357

Qy    278   TTGTAGTGGATGAAAAATCAGGTCAGGTGAGACTCAGGTGACACTGGAACAG    337
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    947   GGAGTTCAGGAGACCTGCTGAGAGCTGTCAGTGACAGTGAGTCCTCAGACATTAGCC    1006

Db    143   TTGTAGTGGATGAAAAATCAGGTCAGGTGACACTGGAAGAATACCAGCACCAGTGGAACAG    202
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1007   AGCCATACGTCGTGTGCCGCAGTGTCCTGAGTACAGAAGGCAGGCGGCGGCAGCCTCCCC    1066

Qy    338   TGATTAACAAGCTGAAGGTTGTTAAGAAGCAGACAGACATGCCCTTTACAGACTGGGGATGTCA    397
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1067   ACTGCCCAGCACCGAGGGCGAGCCAGGGACCCCACAGGCCCTGGGGGATGCACCCTTCCA    1126

Db    203   TGATTAACAAGCTGAAGGTTGTTAAGAAGCAGACAGACATGCCCTTTACAGACTGGGGATGTCA    262
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1127   ACTGCCCAGCACCGAGGGCGAGCCAGAGCCCCACAGCGCCCCACAGGGAGAGTGCACCCTCCA    1186

Qy    398   TCTACTTGGTGTACAGGAGGAGAATGAACCGGAACACAACTGGCATACCTCTATGAAATCTT    457
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1187   CGTCCGTTCAGCCTGACGACAGCAGCAGTCCAGGATTACGTGCCCCTCTGCAAGGAAGCCACG    1246

Db    263   TCTACTTGGTGTACAGGAGGAGAATGAACCGGAACACAACGTGGCATACCTCTATGAAATCTT    322
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1247   CGTCCGTTCAGCCTGACGACAGCAGCAGTCCAGGATTACGTGTGCCCTCTGCAAGGAAGCCACG    1306

Qy    458   TAAGTGAAAAGACAAGGCATGACACAAGAATCCTTTGAAGCTAACAAGGAAAAATGTGTTCC    517
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1307   CCCTGTGCACCTGCTTCCAGCAGTGTCGGGTCTGCCTGCCAGCCTTTCTGCCTGCACCTGTGGG    1366

Db    323   TAAGTGAAAAGACAAGGCATGACACAAGAATCCTT----------    356
            |||||||||||||||||||||||||||||
Qy   1658   GCTGCACCCGGACCGGCTGCTACGGGTCGTGCCTGCCCGTTTTGT    1701

Qy    518   ATGGGACCAAAGATACCTCAGGTGCAGGTCAGGGGCCGATCCCCGGGTCCCTC    577
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1367   GCTGCACCCGGACCGGCTGCTACGGGTCGTGCCTGCCCGTTTTGTGTGTCGGGACCAGAGT    1426

Db    357   --------------------    356
Qy   1702   ----------------GAGCTCAA    1709

Qy    578   CGTCCGTCGCCCGCCACTCAGGTGTCTTTGAGGAACCAACAGACCAATCAACGACGTCAG    637
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1427   GCTTTCCTGCGACATCATTATAAACAGGTAACACATTGCCCCTTCATCTCCGCAGAGCTCAA    1486

Db    357   --------------------    356
Qy   1710   GAAGAATTACCTGCAACCAGAGGTTTGACATGGCAAACATGTTGACCGAGAGCCTCGT    1769

Qy    638   ACCTCTTCCCCACAGCCTCGGCCTCTTCTCCACGGAGCCTTCTCCTGCAGGGCGGAGAGCCGTT    697
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1487   GAAGAATTACCTGCAACCAGAGGTTTGACATGGCAAACATGTTGACCGAGAGCCTCGT    1546

Db    357   --------------------    356
Qy   1770   GGCTCTCCAGCGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTCT    1829

Qy    698   CCTCCAGTTGTGGGGTCTGGGGGTGGGGGCATCTCCCCTAAAGGAAGTGGTCCCTCGTGG    757
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1547   GGCTCTCCAGCGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTCT    1606

Db    357   --------TGGGTGGGGGGTGGGGGGTGGGGGCATCTCCCCTAAAGGAAGTGGTCCCTCGTGG    406
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1830   GTGTTACTGCTGGCCTGCGCAGCTTCCGTGACCTATCAGTATCGGCAGAGAACAT    1889

Qy    758   CAAGTGATGAAGTCTTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAAGACTGCGTCCTTTT    817
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1607   GTGTTACTGCTGGCCTGCGCAGCTTCCGTGACCTATCAGTATCGGCAGAGAACAT    1666

Db    407   CAAGTGATGAAGTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAAGACTGCGTCCTTTT    466
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1667   TCCTGCTTCCGAGTTGCCAGTGGCGTGCTAACATCCCGTCCTGCTACTCGGGCCGTAA    1726

Qy    818   CGTCGTTGGAACCCCAGGATCAGGAGATTTGGAGCCCGTGAAGAAGAAAAATGAGAGGAG    877
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1727   TCCTGCTTCCGAGTTGCCAGTGGCGTGCTAACATCCCGTCCTGCTACTCGGGCCGTAA    1786

Db    467   CGTCGTTGGAACCCCAGGATCAGGAGATTTGGAGCCCGTGAAGAAGAAAAATGAGAGGAG    526
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1787   CTGCCGCACTCAGTGAAAGCTCAGTGAGTTTGCAGCCCTGAGCAGCTTTCAGCAGTGAACAGAC    1846

Qy    878   ATGGGACCTTGACCTGAACGGGCCAGTTGTTGGTCGCCACACCGGCTAGAAATGCCCAAA    937
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1847   CTGCCGCACTCAGTGAAAGCTCAGTGAGTTTGCAGCCCTGAGCAGCTTTCAGCAGTGAACAGAC    1906

Db    527   ATGGGACCTTGACCTGAACGGGCCAGTTGTTGGTCGCCACACCGGCTAGAAATGCCCAAA    586
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1907   AGGGACGCAGGGTCTCCGACAGGTGCTCTGGGGGTGACTCTTCGTGGAGCTTTTTACCCT    1966

Qy    938   CCGTCCACGAGGACGTCAGAGCAGCGGCGGCTCAGAGCCAGCACAGGAGACGCTGA    997
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1967   AGGGACGCAGGGTCTCCGACAGGTGCTCTGGGGGTGACTCTTCGTGGAGCTTTTTACCCT    2069

Db    587   CCGTCCACGAGGACGTCAGAGCAGCGGCGGCTCAGAGCCAGCACAGGAGACGCTGA    646
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2010   CTGCCGCACTCAGTGAAAGCTCAGTGAAAGCTCAATGAATTCAATCATATCTGTGAACAGAC    2069

Qy    998   CATGCATCATCTCCAGGACGTCCGGAGGACCTGCTGCACGACTGCGTGAGTTGCAGCCCTGCATGCACA    1057
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1727   CTGCCGCACTCAGTGAAAGCTCAGTGAAAGCTCAATGAATTCAATCATATCTGTGAACAGAC    1786

Db    647   CATGCATCATCTCCAGGACGTCCGGAGGACCTGCTGCACGACTGCGTGAGTTGCAGCCCTGCATGCACA    706
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2070   AAGGTTCAAAAACTAAGCATCCAGAGGACAAGACAGTCAAGGTGTTTCAGCACTGGGGCCGTAA    2129

Qy   1058   CGTTCTCGCGCGGCTTGCTACTCGGGCTGGATGGAGCGCTCCGTCCCTGTCCCTACCTGCC    1117
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2130   AGCGTGTTTTTAAATACAGAGACAAGACACGTCAAGGTGTTTCACAGCCCCTGAGGGA    2189

Db    707   CGTTCTCGCGCGGCTTGCTACTCGGGCTGGATGGAGCGCTCCGTCCCTGTCCCTACCTGCC    766
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1787   AAGGTTCAAAAACTAAGCATCCAGAGGACAAGACAGTCAAGGTGTTTCAGCACTGGGGCCGTAA    1846

Qy   1118   GCTGTCCGTGAGCCGATCTGTAAAAACCACATCTCAACCTCGTGAAGCATACC    1177
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1847   AGCGTGTTTTTAAATACAGAGACAAGACACGTCAAGGTGTTTCACAGCCCCTGAGGGA    1906

Db    767   GCTGTCCGTGAGCCGATCTGTAAAAACCACATCCAACCTCGTGAAGCATACC    826
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2190   AGGGACGCAGGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCGTGGAGCTTTTTACCCT    2249

Qy   1178   TCATCCAGACATCCAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAGTATGCAGCAGGA    1237
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2250   CTGAGTGAGACCCTCCCCAGAGCCCGCCCGGGGCCCGCCCTCCTCGTGAGCGCTGG    2309

Db    827   TCATCCAGACATCCAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAGTATGCAGCAGGA    886
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1238   ATAAAATCACTCAAGAGACCTGCTGAGAGCTGTCAGCGCCCAAAGTGCAGTCAGGCGGCGTCTTTTTCTGATGAAGAAG    1297
```

```
Db  1967 CTGAGTGAGACCCTCCCAGAGCCCCGGGGGGCGCGAGCCCCGCCCTCCTGGTGAGCGGCTG 2026

Qy  2310 GCAGGGGCTCGTCGTGGCATCAGCAGCAGAGACGAAGCCTTTCTGTAACATGCGGCGCGTC 2369
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2027 GCAGGGGCTCATCGTGGCATCAGCAGCAGAGACGAAGCCTTTCTGTAACATGCGGCGCGTCC 2086

Qy  2370 CGCCGAGAGGGGCAGTTTGCTCTTTTGTACATTTTCGAAACTACAGTTAAAGCACAGAG 2429
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2087 CGCCGAGAGGGGCAGTTTGCTCTTTTGTACATTTTCGAAACTACAGTTAAAGCACAGAAG 2146

Qy  2430 TCTGTTTTTCAGGAAAAGTTTCAAGGGAGAAGGCAAGTTTATCAAAAACATTGTTTCAGG 2489
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db  2147 TCTGTTTTTCAGGAAAAGTTTCAAGGGAGAAGGCAAGTTTATCAAAAACATTGTTTCAGG 2206

Qy  2490 AGAAGGGAGCATAAGTTTACAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAAC 2549
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2207 AGAAGGGAGCATAAGTTTACAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAAC 2266

Qy  2550 CACAGCATTTTATCTATTTTTTATTTTTAAATAGGTTTGGTGCTTATCTTCTAATAAGATTT 2609
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2267 CACAGCATTTTATCTATTTTTTATTTTTAAATAGGTTTGGTGCTTATCTTCTAATAAGATTT 2326

Qy  2610 AAATGTCACAAACTGTAGCACAAAATAATATAATTTATAATTTACAAATTGACAAAAA 2666
              |||||||||||||||||||||||||||||||||||||||||
Db  2327 AAATGTCACAAACTGTAGCACAAAATAATATAATTTATAATTTACAAATTGACTAAAA 2383
```

```
;  RESULT 3
;  US-09-764-864-22
;
;  Sequence 22, Application US/09764864
;  Patent No. US20020132753A1
;  GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE: 2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 22
;  LENGTH: 1311
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: SITE
;  LOCATION: (1281)
;  OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-22

Query Match          35.3%;  Score 945;  DB 9;  Length 1311;
Best Local Similarity 81.1%;  Pred. No. 3.2e-252;
Matches 1267;  Conservative     6;  Mismatches     7;  Indels 282;  Gaps     4;
```

```
Qy    56 TCCGGGGTTCGGCGCCGCCGGGGGGCGCGGGATCGTGAATCCCATGGAGCCGGCCCGAGGAGGCAAGC 115
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5 TCCGGGGTTCGGCGCCGCCGGGGGGCGCGGGGATGTGAATCCCATGGAGCCGGCCCGAGGAGGCAAGC 64

Qy   116 AGTCGGCCGCCGCCGCCGAGCCCTGGGACGGCTCCTGCGTCTGGGCGTCGCGGAGGAGGGGCGAGC 175
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db    65 AGTCGGCCGCCGCCGCCGAGCCCTGGGACGGCTCCTGCGTCTGGGCGTCGCGGAGGAGGGGCGAGC 124

Qy   176 CGCACGTCCTCTGAGGAGGCGGAGTGGACCATCGGGCGACGAGCTGAATTGTAGTTGTGGATGAAAAT 235
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db   125 CGCACGTCCTCTGAGGAGGCGGAGTGGACCATCGGGCGACGAGCTGAATTGTAGTTGTGGATGAAAAT 184

Qy   236 CCTTCCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTGTAGTTAACAAGCTGAAGG 295
              |||||||||||||||||||||||||||||||||||||||||||||
Db   185 CCTTCCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTGTAGTTAACAARCTGAARG 244

Qy   296 CAGGTCAGGTGACACTGGAAGATACCAGCACCAGTGACAGTGATTAACAAGCTGAAGG 355
              |||||||||||||||||||||||||||||||||||||||||||:|
Db   245 CAGGTCAGGTGACACTGGAAGATACCAGCACCAGTGACAGTGATTAACAARCTGAARG 304
```

```
Db   356 TTGTTAAGAAGCAGACATGCCCTTTACGACTGGGGATGTCATCACTTGTGTACAGGA 415
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db   305 TTGTTAAGAARCAGACATGCCCTTTACGACTGGGGATGTCATCACTTGTGTACAGGA 364

Qy   416 AGAATGAACCGGAACAACAACGTGGCATACCTCTATGAATCTTAAGTGAAAGCAAGGCA 475
              |||||||||||||||||||||---
Db   365 AGAATGAACCGGAACACA                                          382

Qy   476 TGACACAAGAATCCTTTGAAGCTAACAAGGAAAAATGTGTTCCATGGGACCAAAGATACCT 535
Db   383 ------------------------------------------------------------ 382

Qy   536 CAGGTGCAGGTGCAGGGGCGAGGGGCCGATCCCCGGTCCCTCCGCCCGTCGCCCGCCACTC 595
Db   383 ------------------------------------------------------------ 382

Qy   596 AGGTGTGCTTTGAGGAACCACCAGCCATCACTGACGTCAGACCTCTTCCCCACAGCCT 655
Db   383 ------------------------------------------------------------ 382

Qy   656 CGGCCCTTCCACGAGGACGCCTTCTCCCTGCAGGGCGGAGAGCGTTCCTCCAGTTGTGGGTCTCG 715
              ||:|||
Db   383 ----------------------------------------GGTCTCG 388

Qy   716 GGGGTGGTGGTGGCATCTCCCCTAAAGGAAGTCGTCCCTCTGTGGCAAGTGATGATGAAGTCTCCA 775
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   389 GGGGTGGTGGTGGCATCTCCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGATGAAGTCTCCA 448

Qy   776 GCTTTGCTCAGCTCTCCCAGACACAGAAAGACTGCGTCCTTTCGTCGTTGGAACCCCAGG 835
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db   449 GCTTTGCTCAGCTCTCCCAGACAGAAAGACTGCGTCCTTTCGTCGTTGGAACCCCAGG 508

Qy   836 ATCAGGAGGATTTGGAGCCCTGAAGAAGAAATGAGAGAGATGGGAGACCTTGACCTGA 895
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   509 ATCAGGAGGATTTGGAGCCCTGAAGAAGAAATGAGAGAGATGGGAGACCTTGACCTGA 568

Qy   896 ACGGGGCAGTTGTTGGTCGCACAACCGTGAAAATGCCCAAACCGTCCACGAGGACGTCA 955
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db   569 ACGGGGCAGTTGTTGGTCGCACAACCGTAGAAAATGCCCAAACCGTCCACGAGGACGTCA 628

Qy   956 GAGCAGCGGCGGGAAGCCAGACAAGATGGAGGACGCTGACATGCATCATCTGCCAGG 1015
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   629 GAGCAGCGGCGGGAAGCCAGACAAGATGGAGGACGCTGACATGCATCATCTGCCAGG 688

Qy  1016 ACCTGCTGCACGACTGCGTGAGTTTGCAGCCCTGAGTTTGCAGCCCTCGCGGCGGCTTGCT 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db   689 ACCTGCTGCACGACTGCGTGAGTTTGCAGCCCTGAGTTTGCAGCCCTCGCGGCGGCTTGCT 748

Qy  1076 ACTCGGGCTGGATGGAGCGGCTCGTCCTGTCCTACCTGTCCCGTGCCCGTGGAGCGGA 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   749 ACTCGGGCTGGATGGAGCGGCTCGTCCTGTCCTACCTGTCCCGTGCCCGTGGAGCGGA 808

Qy  1136 TCTGTAAAAACCAATCCTCAACAACCTCGTGAAGCATACCTCATCCAGCATCCAGACA 1195
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db   809 TCTGTAAAAACCAATCCTCAACAACCTCGTGAAGCATACCTCATCCAGCATCCAGACA 868

Qy  1196 AGAGTCGCAGTGAAGAGATGTGCAAAGTATGATGGCCAAGAATAAAATCACTCAAGACA 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   869 AGAGTCGCAGTGAAGAGATGTGCAAAGTATGATGGCCAAGAATAAAATCACTCAAGACA 928

Qy  1256 TGCTGCAGCCCAAAGTCAACGTTGACAGTGAGTCCTCAGACATTAGCCAGAGGAAGAGGAGTTCAGAGGACCTGC 1315
              ||||||||||||||||||||||||||||||||||||||||||||||
Db   929 TGCTGCAGCCCAAAGTCAACGTTGACAGTGAGTCCTCAGACATTAGCCAGAGGAAGAGGAGTTCAGAGGACCTGC 988

Qy  1316 TGGAGTGTCAGACGTTGACAGTTGTAGTTGGATGATGAAAAT 1375
              |||||||||||||||||||||||||||||||||||||||
Db   989 TGGAGTGTCAGACGTTGACAGTTGTAGTTGGATGATGAAAAT 1048

Qy  1376 GGCAGTGTCCTGAGTACAGAAGGCAGGCGGCGCCTCCCACTGCCCAGCACCCGAGG 1435
              ||||||||||||||||||||||||||||||||||||||||||||
Db  1049 GGCAGTGTCCTGAGTACAGAAGGCAGGCGGCGCCTCCCACTGCCCAGCACCCGAGG 1108
```

```
Qy    851  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  910
Db         CTGCTTCCGAGTTGCCAGTGCCCATCCCCGTCCTGACTGCTACTGGGGCCGTAACT

Qy   2012  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  2071
Db    911  GCCGCACTCAGGTGAAAGCTCACCACGCCATGAAATTCAATCATATCTGTGAACAGAGAA  970

Qy   2072  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  2131
Db    971  GGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGAGGTGAAGAGAG  1030

Qy   2132  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  2191
Db   1031  CGTGTTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTTCACAGCCCCCTGAGGGAAG  1090

Qy   2192  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  2251
Db   1091  GGACGCAGGGTCTCCGACACGGTGCTCTGGGGTGACTCTTCTGTGGGTGACTCTTTCTGTGGAGC  1148

Qy   2252  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  2311
Db   1149  GAGTGAGACCCTCCCCAGAGCCCCCGGGGCCCGCAGCCCCCTGGTGAACATGCGGCCGTCCCG  1208

Qy   2312  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  2371
Db   1209  AGGGCTCGTGGTGGCATCAGCAGCAGAGAGCAGAGAAGCCTTTCTGTAACATGCGGCCGTCCCG  1268

Qy   2372  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  2431
Db   1269  CCGAGAGGGCAGTTTGCTCTTTTGTACATTTTCCGAAACTACAGTAAAGCAGAAGTC  1328

Qy   2432  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  2491
Db   1329  TGTTTTCAGGAAAAGTTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAG  1388

Qy   2492  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  2551
Db   1389  AAGGGAGCATAAGTTACAGCCTACAGCGTACACAATATCCTGCTGCTGGAAAACCA  1448

Qy   2552  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  2611
Db   1449  CAGCATTTTATCTCATTTTTATTTTTAATTAATAGGTTTGGTGCTTATCTTCTTCAATAAGATTTAA  1508

Qy   2612  ||||||||||||||||||||||||||||||||||||||||||||||||||||  2666
Db   1509  ATGTTCACAAACTGTAGCACAAATAATAATAATTTATAATTTACAAATTGACTAAAA  1563
```

```
RESULT 5
US-09-764-864-483
; Sequence 483, Application US/09764864
; Patent No. US2002013753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PT223
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE: 2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 483
;  LENGTH: 693
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: SITE
;  LOCATION: (11)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: SITE
;  LOCATION: (13)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: SITE
```

```
Qy   1436  GCGAGCCAGGAGCCCCACAGGCCCTGGGGGATGACCCTCCAGTCCGTCAGCTGACGA  1495
Db   1109  GCGAGCCAGGAGCCCCACAGGCCCTGGGGGATGACCCTCCAGTCCGTCAGCTGACGA  1168

Qy   1496  CAGCAGTCAGGATTACGTGTGCCCTCTGCAAGGAAGCACGCACCCTGTGCACCTGCTGCT  1555
Db              ||||||||||||||||||||||||||||||||||||||||||||||
           CAGT----CAGGATTACGTGTGCCCTCTGCAAGGAAGCACGCACCCTGTGCACCTGCTGCT  1224

Qy   1556  TCCAGCCCATGCCCGACCGGAGACGCGGAGCGGCAGGA-CCCGCGTGTCGCCCCTCAG  1614
Db   1225  T-CAGCCCATGCCCGACCGGAGAGTGGAACGCGAACAGGACCCGCGTGKCGSCCCTNAA  1283

Qy   1615  CA 1616
Db   1284  CA 1285
```

```
RESULT 4
US-10-094-749-160
; Sequence 160, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
;  APPLICANT: ISOGAI, TAKAO
;  APPLICANT: SUGIYAMA, TOMOYASU
;  APPLICANT: OTSUKI, TETSUJI
;  APPLICANT: WAKAMATSU, AI
;  APPLICANT: SATO, HIROYUKI
;  APPLICANT: ISHII, SHIZUKO
;  APPLICANT: YAMAMOTO, JUN-ICHI
;  APPLICANT: ISONO, YUUKO
;  APPLICANT: HIO, YURI
;  APPLICANT: OTSUKA, KAORU
;  APPLICANT: NAGAI, KEIICHI
;  APPLICANT: IRIE, RYOTARO
;  APPLICANT: TAMECHIKA, ICHIRO
;  APPLICANT: SEKI, NAOHIKO
;  APPLICANT: YOSHIKAWA, TSUTOMU
;  APPLICANT: OTSUKA, MOTOYUKI
;  APPLICANT: NAGAHARI, KENJI
;  APPLICANT: MASUHO, YASUHIKO
;  TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;  FILE REFERENCE: 084335/0160
;  CURRENT APPLICATION NUMBER: US/10/094,749
;  CURRENT FILING DATE: 2002-03-12
;  PRIOR APPLICATION NUMBER: 60/350,435
;  PRIOR FILING DATE: 2002-01-24
;  PRIOR APPLICATION NUMBER: JP 2001-328381
;  PRIOR FILING DATE: 2001-09-14
;  NUMBER OF SEQ ID NOS: 3381
;  SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 160
;  LENGTH: 2186
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-10-094-749-160

Query Match             32.7%;  Score 876.2;  DB 16;  Length 2186;
Best Local Similarity   99.4%;  Pred. No. 7.1e-233;
Matches  890;  Conservative  0;  Mismatches  3;  Indels  2;  Gaps  1;
```

```
Qy   1772  AGAATTACCTGCCAACCAGAGGTTTGACATGAAAAACATGTTGACCGAGAGCCTCGTCG  1831
Db    671  AGAATTACCTGCCAACCAGAGGTTCTCGACATGGAAAAACATGTTGACCGAGAGCCTCGTCGG  730

Qy   1832  CTCTCCAGCCGCGGGAGTGTTTCTGCTCTGTGTCTGTCGTCTGATTACAGAGTCACGGGAGACACCGTTCTGT  1891
Db    731  CTCTCCAGCCGGGAGTGTTTCTGCTCTGTGTCTGTCGTCTGATTACAGAGTCACGGGAGACACCGTTCTGT  790

Qy   1892  GTTACTGCTGTGGCCTGCGCAGCTTCCGTGAGCTGACCTTATCAGTATCGGCAGAACATTC  1951
Db    791  GTTACTGCTGTGGCCTGCGCAGCTTCCGTGAGCTGACCTTATCAGTATCGGCAGAACATTC  850

Qy   1952  CTGCTTCCGAGTTGCCAGTGCCCATCCCCGTAACATCCCCGTCCTGACTGCTACTGGGGCCGTAACT  2011
```

```
; LOCATION: (623)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (640)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (670)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (687)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (690)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-483

Query Match             20.0%;      Score 536.6; DB 9;   Length 693;
Best Local Similarity   92.5%;      Pred. No. 1.6e-138;
Matches      608;  Conservative      7;  Mismatches  13;  Indels  29;  Gaps  4;

Qy    1044 GCCCTGCATGCACACGTTCTGGCGGCTTGCTACTCGGGCTTGGATGGAGCGGCTCGTCCCT 1103
                ||||||||||||||||||| |||||||||||| ||||||||||||||||||||||||||
Db      14 GCCCTGCATGCACACGTTCTGCGCGGCTTGCTCACTCGGGCTGGATGAGCGCGTCGTCCCT 73

Qy    1104 GTGTCCTACCTGCCGCGTCGTCCCGTGGAGGCGGATCGTGAAAAACCACATTCCTCAACAACCT 1163
                |||||||||||||||||| ||||||||||||||||||| ||||||||||||||||||||||||
Db      74 GTGTCCTACCTGCCGCGTCGTCCCGTGGAGCGCGATCGTGAAAAACCACATTCCTCAACAACCT 133

Qy    1164 CGTGGAAGGACATACCTCATCCAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAG 1223
                ||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     134 CGTGGAAGCATACCTCATCCAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAG 193

Qy    1224 TATGGATGCCAGGAATAAAATCACTCAAGACATGCTCAGCCCAAAGTCAGGCGGTCTTT 1283
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     194 TATGGATGCCAGGAATAAAATCACTCAAGACATGCTCAGCCCAAAGTCAGGCGGTCTTT 253

Qy    1284 TTCTGATGAAGAAGGGGAGTTCAGAGGACCTGTCAGAGCTGTCAGAGTCGACAGTGAGTC 1343
                ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     254 TTCTGATGAAGAAGGGGAGTTCAGAGGACCTGTCGAGTCCTGTCAGAGTCGACAGTGAGTC 313

Qy    1344 CTCAGACATTAGCCAGCCATACGTCGTGTGCCCGGCAGTGTCCTGAGTACAGAAGGCAGGC 1403
                |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     314 CTCAGACATTAGCCAGCCATACGTCGTGTGCCCGGCAGTGTCCTGAGTACAGAAGGCAGGC 373

Qy    1404 GGGCAGGCTTCCCCACTGCCCAGCACCGAGGGCGAGCCAGGAGCCCCACAGGCCCTGGG 1463
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     374 GGGCAGGCTCCCCACTGCCCCAGCACCCGAGGGCGAGCCAGGAGCCCCACAGGCCCTGGG 433

Qy    1464 GGATGCACCCTCCACGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTGTGCCCTCT 1523
                ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db     434 GGATGCACCCTCCACGTCCGTCAGCCTGACGACAG----TCAGGATTACGTGTGCCCTCT 489

Qy    1524 GCAAGGAAGCAGGACCACGCCCTGTGCACCTGCTTCCAGCCCATGCCCGACCCGGAGAGCGGA 1583
                ||||||||||||||||||||||||||||| ||||||| |:||||
Db     490 GCAAGGAAGCAGGACCACGCCCTGTGCACCTGCTTC-CAGCCCATGCCCGACCCGGARAGYGGA 548

Qy    1584 GCGGCGGAGCAGGACCCCGGGTCGGGTCAGCCCCTCAGCAGTGTGCGGTCGCTGCCGCAGCCTTTCTG 1643
                ||||||||||||||||||||||||| ||:|:|||
Db     549 -CGCGGARCAGGACCCCGGGTGKCGCGCC-------------TCAMCWTTSTG 584

Qy    1644 CCACCTGTACTGGGGCTGCACCCGGACCCGGCTGCCTACGGCGCCCCGTTTTG 1700
                ||||||||||| |||:|:|||
Db     585 CCACCTGTACTGGGGCTGCACCCCGGACCCGGCTACGGNTTGCTGGCCCCGTTTGNG 641

RESULT 6
US-09-918-995-2180
; Sequence 2180, Application US/09918995
; Publication No. US2003007362A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
```

```
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2180
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(476)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-2180

Query Match             14.7%;      Score 394.4; DB 10;  Length 476;
Best Local Similarity   96.7%;      Pred. No. 5.5e-99;
Matches      413;  Conservative      0;  Mismatches  12;  Indels  2;  Gaps  1;

Qy    2046 ATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCA 2105
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      52 ATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGCCCTGAGCA 111

Qy    2106 GCTTTCAGCACTGAGGTGAAGAGAGGAGAGAGCCGGTGTTTTTAAAATACAGAGACAAGACACGTCAAG 2165
                |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     112 GCTTTCAGCACTGAGGAGGAGGAGAGAGCCGTGTTTTTAAAATACAGAGACAAGACACGTCCAG 171

Qy    2166 GTGTTTTCAGAGCCCCCTGAGGAAGGACACGCAGGGTCTCCGACAGGTGCTCTGGGGTGA 2225
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172 GTGTTTTCAGAGCCCCTGAGGGAAGGACGCAGGGTCTCCGACAGGTGCTCTGGGGTGA 231

Qy    2226 CTCTTCTGTGGAGCTTTTACCCTCTGGGCGGTCGTGGTGGCATCAGCAGGAGACGAAG 2285
                |||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db     232 CTCTTCTGTGGGAGC--TTTACCCTCTGGGCGGTCGTGGTGGCATCAGCAGGAGACGAAT 289

Qy    2286 GCCCGCCCTCCTGGGTGAGCCGTCGGGCAGGGGCCAGTTTTGCTCTTTTGTACATTTT 2345
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     290 GACCCGCCCTCCTGGGTGAGCCGTCGGGCAGGGGCCAGTTTTGCTCTTTTGTACATTTT 349

Qy    2346 CCTTTCTGTAACATGCGGGCCCTCCCGCCGAGGAGGGCCAGTCTGTTTCAGGAAAAGTTTCAAGGAGGAAGGGCAA 2405
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     350 CCTTTCTGTAACATGCGGGCCCTACCCGCCGAGGAGGGCCAGNATTGCTGTTTTGTACATTTT 409

Qy    2406 CCGAAACTACAGTTAAAGCAGAGTCTGTTTCAGGAAAAGTTTCATGAAATAGTTTCAAGGGAGGAAGGGCTA 2465
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     410 CCGAAACTACAGTTAAAGCGGAGTCTGTTTCATGAAATAGTTTCAAGGGAGGAAGGGCTA 469

Qy    2466 GTTTATC 2472
                |||||||
Db     470 GTTTATC 476

RESULT 7
US-09-764-864-322
; Sequence 322, Application US/09764864
; Patent No. US2002013275A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 322
; LENGTH: 357
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
```

```
;   LOCATION: (272)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (300)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-322

Query Match             11.6%;   Score 310.2;  DB 9;  Length 357;
Best Local Similarity   97.7%;   Pred. No. 1.3e-75;
Matches  345;  Conservative   0;  Mismatches   5;  Indels   3;  Gaps    3;

Qy     56 TCCGGGTTCGGCGCCGCGGGGCGGGGATGTGAATCCGGATGGAGCGGCCGAGGAAGGCAAGC 115
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5 TCCGGGTTCGGCGCCGCGGGGCGGGGATGTGAATCCGGATGGAGCGGCCGAGGAAGGCAAGC 64

Qy    116 AGTCGCCGCCGCCGCCAGCCCTGGGACCGGCTCCTGCGCTTCTGGCGCGGAGAGGGGCGAGC 175
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65 AGTCGCCGCCGCCGCCAGCCCTGGGACGGCTCCTGCGCTTCTGGCGCGGAGAGGGGCGAGC 124

Qy    176 CGACGTCCTCCTGAGGAAGCGGGAGTGGACCATCGGGCGGAGAGAGGTTGCGACCTTT 235
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 CGACGTCCTCCTGAGGAAGCGGGAGTGGACCATCGGGCGGAGAGAGGTTGCGACCTTT 184

Qy    236 CCTTCCCCAGCAATAAACTGGTCTCTGAGATCACTGAGATTGTAGTGAGTGGATGAAAAAT 295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    185 CCTTCCCCAGCAATAAACTGGTCTCTGAGATCACTGAGATTGTAGTGAGTGGATGAAAAAT 244

Qy    296 CAGGTCAGGTGACACTGAAGATACCAGCACCAGTGAACAGTGATTAACAAGTGAAGG 355
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    245 CAGGTCAGGTGACACTGAAGATACCANCACCAGTGAACAGTGATTAACAAACTNAAAG 304

Qy    356 TTG-TTAAGAAGCAGACATGCCC-TTTACAGACT-GGGGATGTCATCTACTTG 405
          ||| ||||||||||||||||||| ||||||||| |||||||||||||||||
Db    305 TTGTTTAAGAAACAGACAGCATGCCCTTTTTACAGACTGGGGGATGTCATCTACTTG 357


RESULT 8
US-09-918-995-13662
;   Sequence 13662, Application US/09918995
;   Publication No. US20030073623A1
; GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;   TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;   FILE REFERENCE: 20411-756
;   CURRENT APPLICATION NUMBER: US/09/918,995
;   CURRENT FILING DATE: 2001-07-30
;   PRIOR APPLICATION NUMBER: US/09/235,076
;   PRIOR FILING DATE: 1999-01-20
;   NUMBER OF SEQ ID NOS: 38054
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 13662
;   LENGTH: 449
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(449)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-13662

Query Match             10.2%;   Score 273.6;  DB 10;  Length 449;
Best Local Similarity   98.6%;   Pred. No. 2.5e-65;
Matches  276;  Conservative   0;  Mismatches   4;  Indels   0;  Gaps    0;

Qy   2387 TTGCTCTTTGTACATTTCCGAACTACAGTAAAGCAGAAGTCTGTTTTCAGGAAAG 2446
          || |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     34 TCGCTCTTTGTACATTTCCGAACTACAGTAAAGCGGAAGTCTGTTTTCAGGAATAG 93

Qy   2447 TTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGGAGAAGGGAGCATAAGTT 2506
          ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db     94 TTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGGAGAAGGGGAGCATAAGTT 153
```

```
Qy   2507 TACAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTATCTAT 2566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    154 TACAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTATCTAT 213

Qy   2567 TTTTTATTTTAATAGGTTTGGTGCTTATCTTCAATAAGATTTAAATGTCACAAACTGTA 2626
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    214 TTTTTATTTTAATAGGTTTGGTGCTTATCTTCAATAAGATTTAAATGTCACAAACTGTA 273

Qy   2627 GCACAAATAATAATTTATAATTTACAAATTGACAAAAA 2666
          ||||||||||||||||||||||||||||||||||||
Db    274 GCACAAATAATAATTTATAATTTACAAATTGACTAAAA 313


RESULT 9
US-09-764-864-323
;   Sequence 323, Application US/09764864
;   Patent No. US20020132753A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PT223
;   CURRENT APPLICATION NUMBER: US/09/764,864
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 1792
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 323
;   LENGTH: 354
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (289)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (313)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-323

Query Match              9.6%;   Score 256.4;  DB 9;  Length 354;
Best Local Similarity   93.3%;   Pred. No. 1.3e-60;
Matches  266;  Conservative   1;  Mismatches  18;  Indels   0;  Gaps    0;

Qy    221 GAGGTGCGACCTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTG 280
          ||||| ||||||||||||| |||||||||||||||||| |||||||||||||||
Db     27 GAGATAAGGGACCTGACTTCCCCWGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTG 86

Qy    281 TAGTGGATGAAAAATCAGGTGTTAAGAAGACTGGACACTGGAAGATACCAGCAACAGTGA 340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     87 TAGTGGATGAAAAATCAGGTGTTAAGAAGACTGGACACTGGAAGATACCAGCAACAGTGA 146

Qy    341 TTAACAGCTGAAGGTTGTTAAGAAGCAGAATGCCCTTTACAGACTGGGGATGTCATCT 400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    147 TTAACAGCTGAAGGTTGTTAAGAAGCAGAATGCCCTTTACAGACTGGGGATGTGTCATCT 206

Qy    401 ACTTGGTGTACAGGAAGAATGAACCGGAACAACAGCGGCATACCTCTATGAATCTTTAA 460
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    207 ACTTGGTGTACAGGAAGAATGAACCGGAACAACAGCGGCATACCTCTATGAATCTTTAA 266

Qy    461 GTGAAAGCAAGGCATGACACCAAGAATCCTTTGAGCTAACAAGG 505
          |||||||||||||||||||||||||||||||||||||||||||
Db    267 GTGAAAGCAAGGCATGACACCAANGAATCCTTTGATACCTCAGGG 311


RESULT 10
US-09-764-864-742
;   Sequence 742, Application US/09764864
;   Patent No. US20020132753A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PT223
;   CURRENT APPLICATION NUMBER: US/09/764,864
```

; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 742
; LENGTH: 354
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (289)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (313)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-742

Query Match   9.6%;   Score 256.4;   DB 9;   Length 354;
Best Local Similarity   93.3%;   Pred. No. 1.3e-60;
Matches 266;   Conservative 1;   Mismatches 18;   Indels 0;   Gaps 0;

Qy 221 GAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTG 280
       ||| |||| ||||| ||||||| |||||||||||||||||||||||||||||||||||||
Db 27 GAGATAAGGGACCCTGACTTCCCCWGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTG 86

Qy 281 TAGTGGATGAAAAATCAGGTCAGGTGACACTGGAAGATACCAGCACCAGTGGACAGTGA 340
       ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 87 TAGTTGGATGAAAAATCAGGTCAGGTGACACTGGAAGATACCAGCACCAGTGGACAGTGA 146

Qy 341 TTAACAAGCTGAAGGTTGTTAAGAAGCAGACATGCCCTTTACAGACTGGGGATGTCATCT 400
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 147 TTAACAAGCTGAAGGTTGTTAAGAAGCAGACATGCCCTTTACAGACTGGGGATGTCATCT 206

Qy 401 ACTTGGTGTACAGGAGGAAGAATGAACCGGAACAACCAACGTGGCATACCTCTATGAATCTTTAA 460
       ||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db 207 ACTTGGTGTACAGGAGGAAGAATGAACCGGAACAACAACCAACGTGGCATACCTCTATGAATCTTTAA 266

Qy 461 GTGAAAAGCAAGGCATGACACAACAAGAATCCTTTGAAGCTAACAAGG 505
       |||||||||||||||||||||| |||||||||
Db 267 GTGAAAAGCAAGGCATGACACANGAATCCTTTTGATACCTCAGGG 311

RESULT 11
US-09-908-975-6220
; Sequence 6220, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
SEQ ID NO 6220
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-6220

Query Match   2.2%;   Score 60;   DB 10;   Length 60;
Best Local Similarity   100.0%;   Pred. No. 2.3e-06;
Matches 60;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy 1843 GGAGTGTTTCTGCTGCTGATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGT 1902
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGAGTGTTTCTGCTGCTGATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGT 60

RESULT 12
US-10-424-599-95539
; Sequence 95539, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 95539
; LENGTH: 1255
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1255)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_57283C.1
US-10-424-599-95539

Query Match   1.9%;   Score 51.4;   DB 13;   Length 1255;
Best Local Similarity   57.0%;   Pred. No. 0.0047;
Matches 94;   Conservative 0;   Mismatches 71;   Indels 0;   Gaps 0;

Qy 2515 ACAGGACGTACACAATACCTGCTGCTGGGAAAACCACAGCATTTTATCTATTTTTATT 2574
        || | |||| ||| |||| | ||| || | || ||| | || || ||| |||
Db 858 ACATGTGGTGAACTTTAACTTGGCGTTGTAATAAACGAACCTTTGTATCTCTTATTTTTC 917

Qy 2575 TTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAAT 2634
        || ||| | | |||| ||| ||| ||| | | ||
Db 918 TTGCTAGTAATAGGATGTTTTTACTTCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 977

Qy 2635 AATATAATTTATAATTACAAATTGACAAAAAAAAAAAAAAAAAAA 2679
        |||| || | |||||| ||||| |||||||||||||||||||||
Db 978 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1022

RESULT 13
US-09-814-353-18006
; Sequence 18006, Application US/09814353
; Publication No. US20030165581A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25

```
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 18006
;  LENGTH: 383
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-814-353-18006

Query Match             1.9%;  Score 51;  DB 10;  Length 383;
Best Local Similarity   63.4%;  Pred. No. 0.0027;
Matches   78;  Conservative   0;  Mismatches   45;  Indels   0;  Gaps   0;

Qy      2557 TTTTATCTATTTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAAATGTC 2616
             |||||||||||||||||||||||||| || ||||| |||  |||| |||
Db       116 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAATTAAAAAAAAAAATAAA 175

Qy      2617 ACAAACTGTAGCACAAATAATATATAATTATAATTACAAATTGACAAAAAAAAAAAAAAAA 2676
             | ||| |||||||||||||||||||||||  |||| |||||||||||||||||||||||
Db       176 AAAAAAAAAAAAAAAAAAAAAAAATAAAACAAAAAAAAAAAAAAAAAAAAAAAAA 235

Qy      2677 AAA 2679
             |||
Db       236 AAA 238

RESULT 14
US-10-424-599-127666/c
;  Sequence 127666, Application US/10424599
;  Publication No. US2004031072A1
;  GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT: Zhou Yihua
;  APPLICANT: Kovalic David K
;  APPLICANT: Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE: 2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 127666
;  LENGTH: 411
;  TYPE: DNA
;  ORGANISM: Glycine max
;  FEATURE:
;  OTHER INFORMATION: Clone ID: PAT_MRT3847_86287C.1
US-10-424-599-127666

Query Match             1.9%;  Score 51;  DB 13;  Length 411;
Best Local Similarity   63.4%;  Pred. No. 0.0028;
Matches   78;  Conservative   0;  Mismatches   45;  Indels   0;  Gaps   0;

Qy      2557 TTTTATCTATTTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAAATGTC 2616
             |||||||||||||||||| |||  ||||||||||| |||  |||| |||
Db       216 TCTTATCAATTTTTGTTTTTTGGTTTGTGTGTATTTGTGTTGTTTGTTTTTAACTTTA 157

Qy      2617 ACAAACTGTAGCACAAATAATATATAATTATAATTACAAATTGACAAAAAAAAAAAAAAAA 2676
             | ||| |||||||||||||||||||||||  |||| ||| |||||||||||||||||||
Db       156 AAAAATTTAAAAAAATAAAAAAATAAAAAAATACCAAAAAAAAAAAAAAAAAAAAAAA 97

Qy      2677 AAA 2679
             |||
Db        96 AAA 94

RESULT 15
US-09-960-352-3525/c
;  Sequence 3525, Application US/09960352
;  Patent No. US2002013739A1
;  GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
```

```
;  APPLICANT: Tao, Nengbing
;  APPLICANT: Byatt, John C.
;  APPLICANT: Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE: 2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 3525
;  LENGTH: 434
;  TYPE: DNA
;  ORGANISM: Bos taurus
;  OTHER INFORMATION: Clone ID: 16-BOVMS1-023-Q1-E1-D11
US-09-960-352-3525

Query Match             1.8%;  Score 49;  DB 9;  Length 434;
Best Local Similarity   55.6%;  Pred. No. 0.011;
Matches   94;  Conservative   0;  Mismatches   75;  Indels   0;  Gaps   0;

Qy      2511 GCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTATCTATTTT 2570
             ||||                 ||||||  |||||
Db       211 GCCTTTGGACAGCCCCAATTTCCCAACCTGGGTTTACCCCTGCCCTTTTGGCCCGTT 152

Qy      2571 TATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTGTAGCAC 2630
             |  ||||| | ||| ||||| |||
Db       151 TGGTTTAAGGGTTTTTGCTTTTTGGCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 92

Qy      2631 AAATAATATAATTTATAATTTACAAATTGACAAAAAAAAAAAAAAAAAAAA 2679
             |||  |  |      |      | ||||||||  |||||||||||||
Db        91 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 43

Search completed: May 15, 2004, 00:10:39
Job time : 1598 secs
```

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        May 14, 2004, 12:12:53 ; Search time 10037 Seconds
                                        (without alignments)
                              7970.587 Million cell updates/sec

Title:         US-10-048-046-1
Perfect score: 2679
Sequence:      1 aagaattcggcacgaggccg.........acaaaaaaaaaaaaaaaa 2679

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      2751328 seqs, 1493109276 residues

Total number of hits satisfying chosen parameters:      55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     EST:*
               1:  em_estba.*
               2:  em_esthum:*
               3:  em_estin:*
               4:  em_estmu:*
               5:  em_estov:*
               6:  em_estpl:*
               7:  em_estro:*
               8:  em_htc:*
               9:  gb_est1:*
               10: gb_est2:*
               11: gb_htc:*
               12: gb_est3:*
               13: gb_est4:*
               14: gb_est5:*
               15: em_estfun:*
               16: em_estom:*
               17: em_gss_hum.:*
               18: em_gss_inv:*
               19: em_gss_pln.:*
               20: em_gss_vrt:*
               21: em_gss_fun:*
               22: em_gss_mam:*
               23: em_gss_mus:*
               24: em_gss_pro:*
               25: em_gss_rod:*
               26: em_gss_phg:*
               27: em_gss_vrl:*
               28: gb_gss1:*
               29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                        SUMMARIES
```

Result No.	Score	% Match	Query Length	DB	ID	Description
1	1464.6	54.7	3146	11	AK077629	AK077629 Mus muscu
2	1366.2	51.0	2876	11	AK048930	AK048930 Mus muscu
3	919	34.3	1201	9	AL539798	AL539798 AL539797
4	911.8	34.0	1201	9	AL539797	AL539797 AL539797
5	822.2	30.7	876	13	BQ876252	BQ876252 AGENCOURT
6	774.8	28.9	921	13	BU176559	BU176559 AGENCOURT
7	765.4	28.6	870	13	BQ215731	BQ215731 AGENCOURT
8	758	28.3	992	13	BQ073354	BQ073354 AGENCOURT
9	757.6	28.3	877	14	CD243458	CD243458 AGENCOURT
10	740.6	27.6	910	13	BU148471	BU148471 AGENCOURT
11	727.4	27.2	840	12	BG395714	BG395714 60245836S
12	726.8	26.7	953	13	BQ936072	BQ936072 AGENCOURT
13	714	26.4	938	13	BQ924736	BQ924736 AGENCOURT
14	708.4	26.4	824	12	BG831289	BG831289 602766120
15	706.2	26.4	944	13	BX409790	BX409790 602766120
16	700	26.1	909	13	BX409791	BX409791 BX409791
17	694.4	25.9	729	14	CA916737	CA916737 30LSPRIM2
18	694.4	25.4	1201	9	AL561570	AL561570 30LSPRIM2
19	680	25.4	1116	9	AL528043	AL528043 AL528043
20	674.8	25.2	897	10	BP971929	BP971929 602240332
21	661	24.7	940	13	BQ934867	BQ934867 AGENCOURT
22	652.4	24.4	759	12	BI769973	BI769973 603060267
23	650.4	24.3	916	12	BM459647	BM459647 603060267
24	639.6	23.9	666	10	BE407732	BE407732 601299777
25	637.4	23.8	934	10	BF971838	BF971838 602240424
26	637.2	23.8	1092	12	BM541730	BM541730 AGENCOURT
27	628	23.4	687	9	AU138310	AU138310 AU138310
28	626.8	23.4	690	10	BE280165	BE280165 601158370
29	626.2	23.4	938	13	BQ087242	BQ087242 AGENCOURT
30	624.8	23.3	1022	12	BM558904	BM558904 AGENCOURT
31	624.8	23.0	1096	12	BG335362	BG335362 602404192
32	617	23.0	722	9	AU135780	AU135780 AU135780
33	611	22.8	680	9	AU134937	AU134937 AU134937
34	602.6	22.5	636	10	BE407797	BE407797 601299183
35	588	21.9	635	12	BG481936	BG481936 602526853
36	587.6	21.9	824	13	BQ572677	BQ572677 UI-M-FD0-
37	586.6	21.9	817	13	BM944161	BM944161 UI-M-EHOp
38	583	21.8	824	9	AU132825	AU132825 AU132825
39	578.2	21.6	712	10	BE257482	BE257482 601111439
40	567.8	21.2	650	10	AW957709	AW957709 EST369779
41	567.4	21.2	811	14	CB989147	CB989147 AGENCOURT
42	566.4	21.1	1143	12	BM542545	BM542545 AGENCOURT
43	559.6	20.9	806	14	CB961597	CB961597 AGENCOURT
44	536.2	20.0	982	13	BQ943529	BQ943529 AGENCOURT
45	516	19.3	782	14	CF410232	CF410232 CH3#065_F

```
                        ALIGNMENTS

RESULT 1
AK077629
LOCUS       AK077629        3146 bp    mRNA      linear    HTC  18-SEP-2003
DEFINITION  Mus musculus 8 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:5730484M20 product:similar to CELL CYCLE
            CHECKPOINT PROTEIN CHFR [Homo sapiens], full insert sequence.
ACCESSION   AK077629
VERSION     AK077629.1  GI:26346521
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
```

```
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,K., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
    TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
    TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
    TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3146)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K. Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
    TITLE   Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
  COMMENT   cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3146
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:5730484M20"
                     /db_xref="MGI:2393804"
                     /db_xref="taxon:10090"
                     /clone="5730484M20"
                     /tissue_type="whole body"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="8 days embryo"
     CDS             89..2080
                     /note="unnamed protein product; putative
                     similar to CELL CYCLE CHECKPOINT PROTEIN CHFR [Homo
                     sapiens] (SPTR|Q9NRT4, evidence: FASTY, 81.7%ID,
                     100%length, match=1989)"
                     /codon_start=1
                     /protein_id="BAC36912.1"
                     /db_xref="GI:26346522"
```

```
                     /translation="MELHGEEQPPPQEPWGRLLRLGABEDEPQILLWKREWTIGRRR
                     GCDLSFPSNKLVSGDHCKLTVDEISGEVTLEDTSTNGTVINKLQVVKKQTYPLHSGDI
                     IYLYVRKNEPEHNVAYLYESLSGKQSLTQDSLRANKRNMFHVTKDCSGPGQGDDPQVP
                     LLSPMAQTCLEEPQSTSTSDLLPTASTSSTEPELTSAGQKHSSSSGPGNTSISPKGR
                     SSLVANGELSLSPVFQDKEASFSLLESKDKEAPAKKWKMGDGEDTNLQLLVSGQ
                     RGNAQTSSEDVKDASVKPDKMEBTLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMER
                     SSLCPTCRPVERICKNHILNNLVEAYILQHPDKSRSEDVRSMDARNKITQDMLQPK
                     VRRSFSDEEGSSEDLLELSLVDSESSDISQPYIVCRQCPEYRRQAVQSLPCPVPESL
                     CAVCLQPFCHLYWGCTRGCFCCLAPFCELNLGKCLDGVLNNNYESDILKNYLATR
                     GATLALGGBAPSTSASLPTAPDYMCPLQQSHAICTCFQPMPDRRAEREQDPRVAPQQ
                     GLTWKSVLTESLLALQRGVFMLSDYRITGNTVLCYCCGLRSFRLTYQRQNIPASEL
                     PVTVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQTRFKN"
     polyA_signal    3123..3128
                     /note="putative"
     polyA_site      3146
                     /note="putative"
ORIGIN
```

```
Query Match           54.7%;   Score 1464.6;  DB 11;  Length 3146;
Best Local Similarity 74.5%;   Pred. No. 2.3e-233;
Matches 2007;  Conservative     0; Mismatches 634;  Indels 52;  Gaps 11;
```

```
Qy      18 CCGCAATGTCTCTTGACAGCGGCGGCGGCG------CAGCGCGGTTCCGGTTCGGCGCGG 71
           |||||||||||||||||||||||||||||||        |||||||||||||||||||
Db      10 CCCAGATGTCTGTTGACAGCGGTGGCGGCGGCGACTGTAAGGCGCGGTTCCGGGATCCGGCAG 69

Qy      72 GGGGGGGATGTGAATCCGATGGAGGGCGCCGAGGAAGGGCAAGCAGTCGCCGCGCCCGCA 131
           |||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db      70 GGCGTGGGATGTGAGTCTATGGAGCTACACGGGGAAGAGCAGCCGCCGCCGCGCCGCAGGA 129

Qy     132 GCCCTCGGGGACGGCTCCTGGCGTCTGGGCGGCGGAGGGGCGGAGCGCACGTCTCCTGAG 191
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130 ACCCTCGGGGAGGCTTCTTCGTCTCAGGCGCAGAGGAGTTGCGACCTTTCCTTCCCCAGCAATGA 189

Qy     192 GAAGCGGGAGTGGACCATCGGGCGGAGGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAA 251
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     190 GAAACCGCAGTGGACCATCGAAGGAGGAGGAGGTGTGACCTCTCTTTCCCCAGCAATAA 249

Qy     252 ACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGTCAGTGACACT 311
           |||||||||||||||||||||||||||||||||||||||||  |||||||||||||
Db     250 ACTGGTCTCTGGAGATCACTGTAACTTACAGTGATGATGATAATATCTGGTGAGTGACACT 309

Qy     312 GGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTAAGAGCAGAC 371
           |||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db     310 GGAAGACACCAGCACCAATGGAACAGTGATCAATAAGCTGCAAGTTGTTAAGAGAGCAGAC 369

Qy     372 ATGCCCTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAATGAACCGGAACA 431
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db     370 TTACCCTTACATAGCGGGATATCATCATTGGTGTACAGGAAGAATGAGCCAGAACA 429

Qy     432 CAACGTGGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGGCATGACAACAAGAATCCTT 491
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     430 CAATGTGGCATACCTCTATGAATCTTTAAGTGGCAAACAGAGCTTAACTCAAGACTCCCT 489

Qy     492 TGAAGCTAACAAGGAAAATGTGTTCCATGGACCACCAAAAGATACCTCAGGTGCAGGTGCAGG 551
           ||||||||||||||||||||||||||||||||||||||||||     |||||||
Db     490 TGAAGCCAATAAGGAAAATATGTTCCATGTGACCAAAGGATTGCT------CAGTCCAGG 543

Qy     552 GCGAGGGGCCGATCCCGGGTCCTCCGGTTCCACTTCAGGTGTGCTTTGACGA 611
           |||||||||||||||||||||||||||||||||||||  ||||||||
Db     544 GCAGGGTGATGATCCCAGGTTCCACTATGTCACCCATGGCTCAGACATGCTTAGAGGA 603

Qy     612 ACCACAGCCATCAACATGACGTCAGACCTCTTCCCCACAGCCTTCCCCACGGA 671
           ||||||||||||||||||||||||||||||     |||||
Db     604 ACCACAGCCATCAACATGACATCAGACCTCTTCCCCACGGGCCTCTACCTCTTCTACGGA 663

Qy     672 GCC------TTCTCCTGCAGGGCGGAGGCGTTCCTGAGCCTTCCCAGTGTGGGTCTGGGGTGGTGG 725
           ||||||                |||||||||||||||||||
Db     664 GCCAGAGCTGACCCTGCAGGGCAAAGCATTCTTCTAGCTCTGGACCTGGGAACACAAG 723

Qy     726 CATCTCCCCTAAAGGAAGTGGTCCCTGTCTGGCAAGTGATGAAGTCTCCAGCTTTGCCTC 785
           ||||||||||   |||||||||||||||||||||||||||||||||||||||||||
Db     724 CATCTCCCCAAAAGGACGCAGTCACTTGTTGCAAATGGCAAATGGCGAACTCTCTAGCCTTTCTCC 783
```

```
Qy   1866  CAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGCTGGCCTGCTGGGCTGCGCAGCTTCCGTGAGCT  1925
           ||||||||   || |||||  |||||||||||||  |||  |||||||| ||||||||||||||||||||
Db   1861  CAGAATCACTGAAATACTGTGTCGTGTGTTACTGCTGCTGCTGGTCGCTGGTCGCGTAGCTTCCGAGAGCT  1920

Qy   1926  GACTTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTGAGTTGCCAGTCAGTGCCGTAACATCCCG  1985
           |||||||||||| ||||||||||||||||||||| |||||||||||||||||||||||||
Db   1921  GACCTACCAGTATCGTCAGAACATTCCTGCTTCTGAGTTGCCAGTGACTGCCAGTGCCGTAACATCCCG  1980

Qy   1986  TCCTGACTGCTACTGGGCGCTAACTGCCGCCACTCAGGTGAAAGCTCACCACGCCATGAA  2045
           |||||||||||||||| |||  |||  |||| ||||||||||| |||||||||||||||||
Db   1981  TCCTGACTGCTACTGGGCGCTAACTGTCGCCACTCAGGTGAAGGCTCACCATGCAATGAA  2040

Qy   2046  ATTCAATCATATCTGTGAACAGACAAGGTTCAAAACTAAGCATCCAGAGAGCCCTGAGCA  2105
           ||||||||||||||||||||||||||||||||||  ||| |||||||||||||||||||
Db   2041  ATTCAATCACATCTGTGAGCAAACAAGGTTCAAGACACGTTGAAGACAAGCACGTCAAG  2100

Qy   2106  GCTTTCAGCACTGGAGTGAGAGAGCGTGTTTTAAAATACAGAGACAAGCACGTCAAG  2165
           |||||||||| |||||||||| |||  |||||||||  |||||
Db   2101  AACAGTTACAGAGCA-CAGGAGGCAATGTTTTTAAAAGGGCCCATCAAACAACTTGAAG  2159

Qy   2166  GTGTTTTCACAGCCCCCTGAGGGAAGGACGCCAGGGTCTCCGACAGTGCTCTGGGGTGA  2225
           || |||||||||||| |||||| |||||||  |||||||||||||||||||||| |||
Db   2160  GGCATTTTTACAGTTTTCCCTACGCAGCAGTGGGGGTCTGGTCAGGCACTGGGGTGTGG  2219

Qy   2226  CTCTTCTGTGGAGCTTTTTTACCCTCTGAGTGAGCGCGTCCCCGGGGGGCCGCA  2285
           |||||||||||||| ||||||  ||| ||||||||||||| |||||||||||||| ||
Db   2220  ATTTTCCCTTTGTGGAGCGTCATCGTCAGTGACACACCACGTGCCCCAGTAAC-CT  2278

Qy   2286  GCCCGCCCCTGGTGAGCGCGTCGGGCAGACATGCTCAGCAGCGAG------  2339
           || |||||  ||| |||||| || ||||  |||||||||||||||||
Db   2279  GGGAGGGCCTCCTGGCCACACCAGCCTTCTAGTCGGTCGGTTGAAGAGGATCAGTAGACTGAGTT  2338

Qy   2340  --------ACGAAGCCTTTCTGTAACATGCGACCGTCCGCCGAGAGGGGCAGTT  2386
                   ||||  |||  |||||||| |||||| |||||||||||||||||| |
Db   2339  CTCCCTCCCTCCCCACTGACCGCTCTCCGTCGTGTCTGAGGCAGTGCAGTGCAGGT  2398

Qy   2387  TTGCTCTTTTGTACATTTCCGAAACTACAGTTAAAGCACAGAGTCTGTTTCAGGAAAAG  2446
           ||||||||||| |||||     ||| |||||  ||||||||
Db   2399  TTGTACTTTTGTATCTTTT------CTACCTTGTACCTTGACTATGGTTAAAATGGGGAG  2452

Qy   2447  TTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTCAGGAG--AAGGGAGCATAAG  2504
           ||||||||||||||  ||  |||||||  ||||| |||||||||||  ||||||||||||
Db   2453  CTCATAGGAAAAGTTGTCAGCTGTGTCAGACGCATTCTCCGGAAGGAAGGAGCATACG  2512

Qy   2505  TTTACAGCCTACAGGACGT-ACACAATATCCTGCTGCTGTGGGAAAACCACAGCATTTATC  2563
           ||||||||||||||||||  |  || ||||||| |||||||||||||||| |||||||| |
Db   2513  TTTACAGCCTCCAGGGGTACATAGGATAGTCTCCTCTCTGGAAAGACACAGCCTTTTGTA  2572

Qy   2564  TATTTTTATTTTAATAGGTTTGGTGCTTA-TCTTCTAATAAGATTTAAATGTCACAAAC  2622
           ||||| ||||||||||||||||||||||||      ||||| |||||||||||| ||||
Db   2573  TATTTTTATTTTAATAGGTTTGGTGCTTAGTCGTCCTTAGTCTTGTAATAAGATTTAAATATTTAAAAAC  2632

Qy   2623  TGTAGCACAAATAATATATATTACAAATTGACAAAAAAAAAAAAAAA  2675
           |||||||||||| |||| |||| |||||| |||||||||
Db   2633  TGTAGCACAAATACTATATAATTTATAATTTACAAATTAGACTAAAATTAGGAATA  2685
```

RESULT 2

LOCUS   AK048930   2876 bp   mRNA   linear   HTC 20-SEP-2003

DEFINITION Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length
            enriched library, clone:C230082M18 product:similar to CELL CYCLE
            CHECKPOINT PROTEIN CHFR [Homo sapiens], full insert sequence.

ACCESSION   AK048930

VERSION   AK048930.1   GI:26093209

KEYWORDS   HTC; CAP trapper.

SOURCE   Mus musculus (house mouse)

  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1

AUTHORS   Carninci,P. and Hayashizaki,Y.

TITLE   High-efficiency full-length cDNA cloning

JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
MEDLINE   99279253
PUBMED   10349636
REFERENCE 2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
           Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
           prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED   11042159
REFERENCE 3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
           Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
           Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
           Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
           Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
           Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
           Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
           sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE 4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
           FANTOM Consortium.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
           Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
           of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2876)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
           Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
           Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
           Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
           Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
           Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
           Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
           Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
           Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
           Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
           Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
           Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
           Physical and Chemical Research (RIKEN), Laboratory for Genome
           Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
           RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
           Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
           URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
           Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
           Encyclopedia Project of Genome Exploration Research Group in Riken
           Genomic Sciences Center and Genome Science Laboratory in RIKEN.
           Division of Experimental Animal Research in Riken contributed to
           prepare mouse tissues.
           Please visit our web site for further details.
           URL:http://genome.gsc.riken.go.jp/
           URL:http://fantom.gsc.riken.go.jp/.
FEATURES   Location/Qualifiers
     source   1..2876
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:C230082M18"
                     /db_xref="MGI:2416097"
                     /db_xref="taxon:10090"
                     /clone="C230082M18"

                     /tissue_type="cerebellum"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="0 day neonate"
     misc_feature   1..2876
                     /note="similar to CELL CYCLE CHECKPOINT PROTEIN CHFR [Homo
                     sapiens] (SPTR|Q9NRT4, evidence: FASTY, 81.7%ID,
                     100%length, match=1989)"

ORIGIN

Query Match   51.0%;   Score 1366.2; DB 11; Length 2876;
Best Local Similarity 74.7%; Pred. No. 5e-217;
Matches 1862; Conservative 0; Mismatches 583; Indels 46; Gaps 10;

```
Qy    214 CGGAGACGAGGTTGCGACCTTTCCTCCCCAGCAATAAACTGGTCTCTGGAGATCACTGT 273
              ||||||||||||| |||| ||||||||||||||||||||| |||||||||||||||||
Db    390 CTGAAAAAGGCTGTGACCTCTCTTTCCCCAGCAATAAACTGGTCTCTGGAGATCACTGT 449

Qy    274 AGAATTGTAGTGGATGAAAAATCAGGTCAGGTGACACTGGAAGATACCAGCAGTGGA 333
              | | |||||| |||||||| ||||||||| ||||| |||||||| ||||| |||||
Db    450 AAACTTACAGTGGATGAAATATCTGTGAGTGAGCTGGAAGAACACCAGCACCAATGGA 509

Qy    334 ACAGTGATTAACAAGCTGAAGGTTGTTAAGAACAGACATGCCCTTTACAGACTGGGGAT 393
              |||||||||| ||||||||| |||||||||||||||||||  | |||||||||| |||
Db    510 ACAGTGATCAATAAGCTGCAGTTGTTAAGAACAGAGACTTACCCTTTACAGAGCGGGGAT 569

Qy    394 GTCATCTACTTGGTGTACAGGAAGAATGAACCGGAACACAACGTGCATACCTCTATGAA 453
              || |  |||| |||||| ||||||||||||| |||||||||||| ||| ||||| |||||
Db    570 ATCATCTATTGGTGTACAGGAAGAATGAGCTCAGGAACACAATGTCGCATACCTCTATGAA 629

Qy    454 TCTTTAAGTGAAAAGCAAGGCATGACACAAGAATCCTTTGAAGCTAACAGGGAAAATGTG 513
              ||||||||||| ||||| ||| ||||| || |||||| ||| | ||||| ||| ||| |
Db    630 TCTTTAAGTGGCAAACAGAGCTTAACTCAAGACTCCCTTGAAGCCAATAAGGAAAAATATG 689

Qy    514 TTCCATGGGACCAAAGATACCTCAGGTGCAGGTGCCAGGTGCAGGGCCGATCCCGGGTC 573
              |||| ||||||||| | |  |||| ||||     ||||||||||||| |||||||||||
Db    690 TTCCATGTGACCAAAGATTGCTCAGGTCCAG------GGCAGGGTGATGATCCCCAGGTT 743

Qy    574 CCTCCGTGCGCCCGCCCACTCAGGTGTCTTGAGGACCAGCCATCAACATCGACG 633
              |||||||||| ||||||||||||||||| || |||||||||||| ||||||||||
Db    744 CCACTATTGTCACCCATGGCTCAGACATGCTTAGAGGAACACAGCCATCAACATCGACA 803

Qy    634 TCAGAACCTCTTCCCCACAGCCTCGGCTTCCTCCACGGAGCC-------TTCTCCTGCAGGG 687
              ||||||||| ||||| ||||| ||||| |||| ||| ||||||       |||||||||||||
Db    804 TCAGAACCTCCTCCCACGGCCTCTACCTCTTCTCTCACGGAGCCCAGAGTCTGCACCTCTGCAGGG 863

Qy    688 CGGAGAGCGGTTCCTCCAGTTGTGGGTCTGGGGTCGGAGGTCGTGGCCATCTCCCTAAAGGAAGTGGT 747
              ||||||||||||||| |||||||||| |||| |||||||||||||| |||||| ||||||| ||| ||
Db    864 CAAAAGGCATTCTTCTAGCTCTGGACCTGGGGGAACAACAAGCATCTCCCCAAAAGGACGCGCAGT 923

Qy    748 CCCTCGTGGCAAGTGATGAAGTCTTCAGCTTTGCCTCAGCTCTCCCAGACCAGAAAGACT 807
              |||||||| ||| ||||| |||||||| |  ||||| ||  |||| |||||||||||| |
Db    924 TCACTTGTTGCAAATGGCAAATGGCAACTCTCTAGCCTTTCTCCAGTTTTCCAAGACAAGA--A 980

Qy    808 GGGTCCTTTTCGTGCGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCCCGTGAAGAAGAAA 867
              || |||||||||| ||| |||| ||| ||||| ||  ||||||||||||||  || |||||||
Db    981 GCATCCTTTTCTTTGCTGGAAAGTAAAGAACCATGAGGAATTGGAGCCTGCCAAAAAAAAAA 1040

Qy    868 -ATGAGGAGGATGGGGACCTTGACCTGAACGGGCAGTTGTTGGTCGCACAACCGCGTAG 926
               ||||| |||| ||||||| |||||||||||||  |||  | |||||| |||||||||||
Db   1041 GATGAAAGGAGATGGGGAACTTGACACGAACCTCCAGTTATTAGTTTCAGGCGGCCAGCGTGG 1100

Qy    927 AAATGCCAAACCGTCACGAGACGTCAGAGCACTGAGAGCGGCTGGGAAGCCAGCAAGAATGA 986
              |||||||||||| |||||||| ||||||||  |||||||||||| |||||||||||  ||||
Db   1101 AAATGCCAAACCTCAAGTCAAGATGTCAAAGATGCCTCTGTGAAGCCAGCAAGAAGCAAGATGA 1160

Qy    987 GGAGACGGCTGACATGCATCATCTGCCAGAACCTGCTCACGACTGCGTGGAGTTTGCAGCC 1046
              ||||| ||||||  ||| ||| ||||||| ||| ||||  ||||  ||| ||||| |||||
Db   1161 GGAGACACTAACCTGTATCATCTGCCAGAACCTTCTGCAGACCTTCTGCGACGATTGTGTGAGTTTGCGGCC 1220

Qy   1047 CTGCATGCACACGTTCTGCCGGCCGGTGCTACTCGGGTCGGATGGAGGCCGCTCGTCCGCCTGTG 1106
              |||| ||| ||| ||||| ||||| |||| ||  ||||||||  ||||||||||||| |||||||
Db   1221 TTGTATGCACACATTTTGTCGGCGGTTGCTACTCGGTTGGATGGAGGAGCCGTTCATCTGCTGTG 1280
```

```
Db  2360 ACGGAGCACTGGGGTTCTGGTCAGGCACTGGGGTGTGGATTTCCCTTTGTGGAGGGTC 2419

Qy  2247 CCTCGAGTGAGACCCTCCCAGAGCCCGGGGGCCCCAGCCCGCCCCTCGTGAGCGC 2306

Db  2420 ATCGTCAGTGACACACACCACGTGCCCAGTAAC-CTGGGAGGGCCTCCTGCCACACC 2478

Qy  2307 TGGGCAGGGCTCGTGGTGGCATCAGCAGCAGA----------------GACGAAGCC 2347

Db  2479 AGCCTTCAGTCGGTGTAAGAGGATCAGTAGACTGAGTCTCCCTCCTCCCACTGGACC 2538

Qy  2348 TTTCTGTAACATGCGGCCGTCCCGCCGAGAGGGCAGTTTGCTCTTTTGTACATTTTCC 2407

Db  2539 GCTCTCCCTCCTGTGTCTGAGGCAGTGACAGTGCAGGTTTGTGTATCTTTT-- 2596

Qy  2408 GAAACTACAGTTAAAGCAGAAGTCTGTTTCAGGAAAAGTTTCAGGAAAAGTTGTCAGT 2467

Db  2597 ----CTACCTTGTACCTTGACTATGGTTAAAATGGGAAGCTCATAGGAAAAGTTGTCAGC 2652

Qy  2468 TTATCAAAAACATTGTTTCAGGAG--AAGGGAGCATAAGTTTACAGCCTACAGGACG-TA 2524

Db  2653 TGTGTCAGACGGCATTCTCCGGAGGAAGGAAGGAAGCATACGTTTACAGCCTCCAGGGGTACA 2712

Qy  2525 CACAATATCCTGCTGCTGGGAAAACCCAGCATTTTATCTATTTTTTTTTTATTTTAATAGGTT 2584

Db  2713 TAGGATAGTCTCCTGCTCGGAAAGACACAGCCTTTGTATATTTTTATTTTAATAAGTT 2772

Qy  2585 TGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATATAAATT 2644

Db  2773 TGGTGCTTATCTTCTAATAAGATTTAAATATTAAAACTGTAGCACAAATACTATAATTT 2832

Qy  2645 ATAATTTACAAATTGACAAAAAAAAAAAAAAA 2675

Db  2833 ATAATTTACAAAATTGACTAAAATTAGGAATA 2863
```

RESULT 3
AL539798

LOCUS   AL539798   1201 bp   mRNA   linear   EST 31-MAY-2003
DEFINITION AL539798 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF031YA23 5-PRIME, mRNA sequence.
ACCESSION   AL539798
VERSION   AL539798.2 GI:31264361
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE   Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT   On Feb 15, 2001 this sequence version replaced gi:12869349.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr; Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6792.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF031AA12QP1&cluster=6792.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fullength.invitrogen.com/ InvitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF031AA12QP1.
FEATURES   Location/Qualifiers
     source   1..1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF031YA23"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"

```
Qy  1107 TCCTACCTGCCGCTGTCCCGTGAGCGGATCTGTAAAAACCACATCCTCAACAACCTCGT 1166

Db  1281 CCCTACCTGCCGATGTCCCAGTGGACGGGATTTGCAAAAACCACATCCTGAACAACCTAGT 1340

Qy  1167 GGAGCATATCTCCAGCATCATCCAGACAAGAGTCGCAGTGAAGAGATGTGCAAAGTAT 1226

Db  1341 GGAAGCATACCTTTATCCAGCACCCAGATAAAAGTCGCAGTGAAGAAGATGTGAGAAGTAT 1400

Qy  1227 GGATGCCAGGAATAAAAATCACTCAAGACAATGCTGCAGCCCAAAGTCAGGCGGGCTTTTTC 1286

Db  1401 GGATGCAAGAATAAAAATCACTCAAGATATGCTGCAACCCAAAGTCAGGAGGTCTTTCTC 1460

Qy  1287 TGATGAAGAAGGGAGTTCAGAGGACCTGCTGGAGCGTGCAGCGTTGACAGTGAGTCGTC 1346

Db  1461 TGATGAAGAAGGGGAGTTCAGAGGACCTGCTGGAGCGTGCTGCAGTGACAGTGAATCCTC 1520

Qy  1347 AGACATTAGCCAGCCATACGTCGTGCCCAGTGCCCAGTGTACAGAGAGGCAGGCGGC 1406

Db  1521 AGATATCAGTCAGCCATACATTGTCTGCAGACACCCCGAGTGCCCAGGCGCAGCGAGT 1580

Qy  1407 GCAGCCTCCCCACTGCCCCAGCACCCGAGGGCGAGCCAGAGCCCCAGGCCCTGGGGGA 1466

Db  1581 GCAGTTCTTTCCTTGCCCAGTCCCAGTGGGAGCTACACTGGGCCTCGGCCCTTGGTGG 1640

Qy  1467 TGCACCCTCCAGTCGTCGTCGTGACACAGCAGTCTCAGGATTACGTGTGCCCTCGCA 1526

Db  1641 GGAGGCACCTTGCCAGTTGCCAACAGCCCCGATTACATGTGCCCCTCTTCA 1700

Qy  1527 AGGAAGCCACGCCCTGTCGTGCCATGCCCAGCCCATGGCCGAGCGGGAGCG 1586

Db  1701 AGGAGGCCATGCCATATGCCCGCTGTCCCAGCCATGTGCCTGACCGAGAGCTGAACG 1760

Qy  1587 CGGAGCAGGACCCGCGTGTCGCCCCTCAGCAGTGTGCGGTCTGCCTGCAGCCTTTCTGCCA 1646

Db  1761 GGGAGCAGGAGATCCCCGCGTCCCCGCCCCCTCAGCAGTGTGCGGTGTCGCAGCCCTTCTGCCA 1820

Qy  1647 CCTGTACTGGGGCTGCACACCGGACCGGACCGGCTGCCTGCTGGGGGCCCGTTTGTGAGCT 1706

Db  1821 CCTGTACTGGGGCTGCACAGAGGACTGGCTGCTGCTGGGGCCCCATTCTGTGAGCT 1880

Qy  1707 CAACCTGGTGACAAGTGTCTGGACCGGCGTGCTGAACAACAACAGCTACGAGTCAGACAT 1766

Db  1881 CAACCTGGGGGACAAGTGCTTGGACGTCTGAACAATAACAATTAGAATCGGACAT 1940

Qy  1767 CCTGAAGAATTACTGGCCAACAGAGGTTTGACATGGAAAAAACATGTTGACCGAGAGCCT 1826

Db  1941 CCTGAAGAATTACTGGCCAACCAGGGGTCTGACATGGAAAAGTGTGTTGACAGAGAGTCT 2000

Qy  1827 CGTGGCTCTCCAGCGGGAGTGTTCTGCTCTGATTCAGAGCTGCACGGGAGACACCGT 1886

Db  2001 CCTGGCTCTCGCAGCGAGGTGTTATTGCTCTGATTACGAGATCTGGGAAAATACTCGT 2060

Qy  1887 TCTGTGTTACTGCTGTGGCCTGCGCAGCTTCCGTGAGTCGTGTGACCCTATCGGCAGAA 1946

Db  2061 GCTGTGTTACTGCTGTGGCCTGCGTAGCTTCCGAGAGCTGACCTACCAGTATCGTCAGAA 2120

Qy  1947 CATTCTGCTTCCGAGTTGCCGTAACATCCCGTCCTGACATCCTACTGGGGCCG 2006

Db  2121 CATTCCTGCTTCTGATTGCCAGTGACTGTAACATCCCGTCCTGACTGGGGGCCG 2180

Qy  2007 TAACTGCCGCACTCAGGTGAAAGCTCACCACGCCATGAAAATTCAATCATATCGTGAACA 2066

Db  2181 TAACTGCCGCACTCAGGTGAAGGCTCACCACATGAAAATTCAATCATACACTCTGTGAGCA 2240

Qy  2067 GACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAGGTGAA 2126

Db  2241 AACAAGGTTCAAGAACTGAATATTGGAAGAGACACGTGAAACAGTTACAGAGCA-CAGGA 2299

Qy  2127 GAGAGCGTGTTTTTAAAATACAGAGACAAGCAAGACGTTCACAGCCCCCTGAG 2186

Db  2300 GGGCAATGTTTTTTAAAAGGCCCCATACAAACACTTGAAGGGCATTTTTACAGTTTTCCT 2359

Qy  2187 GGAAGGGACCGCAGGGTCTCGGCACAGGTGCTCTCGGGGTGACTCCTTCTGTGGGAGCTTTTAC 2246
```

```
Db    966 GCGGCTGGGAGCGAGCAAGAGAMAGATGGAGGRG-MSCTGACATGCATM--TCTGCAAGACCTG 1022
              ||||||||  |||:||||||  :::||||||||||  ::||||||||||   || ||||||||||
Qy   1021 CTGCACGACTCCGTGAGTTTGCAGCCCTGCA 1051
              |||::||:        ||||||||||||||||
Db   1023 CTGCMCRMTGCTGAGTTTSAGCCYGCATGCM 1053


RESULT 4
AL539797/c
LOCUS       AL539797 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION  CS0DF031YA23 3-PRIME, mRNA sequence.
ACCESSION   AL539797
VERSION     AL539797.2  GI:31264360
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12869347.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web.: www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to cluster 6792.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF031A12NP1&cluster=6792.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF031A12NP1.
FEATURES             Location/Qualifiers
     source          1..1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF031YA23"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

Query Match       34.0%;  Score 911.8;  DB 9;  Length 1201;
Best Local Similarity   91.2%;  Pred. No. 1.8e-141;
Matches  973;  Conservative  41; Mismatches  47;  Indels  6;  Gaps  5;

Qy   1580 CGGAGCCGGAGCGAGGACACCCGCGTGTCCGCCCCTCAGCAGTGTGCGGGTCTGCCCTGCAGCGTT 1639
              |||||||||||||  || |::|||||||  |||:||||||  ||| |:|||||
Db   1061 CGGRGGGGGGGGCGGGRSMAGACCCGCGTGTCSCCCCTCAGCAGKGCGGKGCGGTGCTGCCAGCCT 1002

Qy   1640 TCTGCCACCTGTACTGGGGTCGAGGTCGACCAGTGTCTGGACGGCGTGCTGAACCAGCTACCAGGAGT 1699
              ||||||||||||:   |||||:||||| ||| |:||||||||   |:||||||||||||  ||| |||
Db   1001 TCTGCCACCTGTACT-GGGCTGCACCCGGACCCGCTG-YACCGGCTGGCCTGCTGGGCCCG-TTT 945

Qy   1700 GTGAGCTCAACCTGGGTGACAGTGTCTGGACGGCGTGCTGAACAACAGCTACGAGT 1759
              |:||||||||||||  |  ||||:||||| ||  | |:||||||||||||||||||
Db    944 GKGRGCTCAACCTGGGTGACAGTGTCGACACGGCGTGCTGAACAACAGCTACGAGT 885

Qy   1760 CAGACATCCTGAAGAATTACCTGGCAACCAGAGGTTTGACATGCATGCATGAAAAACATGTTGACCG 1819
              ||||||||||||||||||||||||:|||||||||||||:||||||||||||||||||||||||||||
Db    884 CAGACATCCTGAAGAATTACCTGGCAACCAGAGGTTTGACATGAAAAACATGTTGACCG 825
```

```
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."

Query Match       34.3%;  Score 919;  DB 9;  Length 1201;
Best Local Similarity   96.8%;  Pred. No. 1.1e-142;
Matches  959;  Conservative  9; Mismatches  19;  Indels  4;  Gaps  3;

Qy     61 GTTCGGCGCGGGGGCGGGGGGATGTGAATCCCGATGGAGCGGCCCGAGGAAGGCAAGCAGTCG 120
              |||||||:|||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db     67 GTTCGGCGCGGGGGCGGGGGGATGTGAATCCCGATGGAGCGGCCCGAGGAAGGCAAGCAGTCG 126

Qy    121 CCGCGCCCGCAGCCCTGGGGACGGCTCCTGCGTCTGGGCGCTGGGCGCTGGGAGGGCGAGCCCGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    127 CCGCGCCCGCAGCCCTGGGGACGGCTCCTGCGTCTGGGCGCTGGGCGCTGGGAGGGCGAGCCCGAC 186

Qy    181 GTCCTCCTGAGGAGCGGGAGTGGACCATCGGCGGCGGAGACGAGGTTGCGACCTTTCCTTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 GTCCTCCTGAGGAGCGGGAGTGGACCATCGGCGGCGGAGACGAGGTTGCGACCTTTCCTTC 246

Qy    241 CCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247 CCCAGCAATAAACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGT 306

Qy    301 CAGGTGACACTGGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGTCGAAGGTTGTT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 CAGGTGACACTGGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGTCGAAGGTTGTT 366

Qy    361 AAGAAGCAGACACATGCCCTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    367 AAGAAGCAGACACATGCCCTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAAT 426

Qy    421 GAACCGGACACAACGTGGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGCAAGCATGACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    427 GAACCGGACACAACGTGGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGCAAGCATGACA 486

Qy    481 CAAGAATCCTTGAAGCTAACAAGGAAAATGTGTTCCATGGGACCCAAAGATACCTCCAGGT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    487 CAAGAATCCTTGAAGCTAACAAGGAAAATGTGTTCCATGGGACCCAAAGATACCTCCAGGT 546

Qy    541 GCAGGTGCAGGGGCAGGGCGAGGGGGCGATCCCCGGGTCTCGTCGCCCCGCCCACTCAGGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    547 GCAGGTGCAGGGGCAGGGCGAGGGGGCGATCCCCGGGTCCCCGTCGCCCCGCCCACTCAGGTG 606

Qy    601 TGCTTTGAGGAACCACCACAGCCATCAACATCGACGTCAGAACCTCTTCCCACAGCCTCGGCC 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    607 TGCTTTGAGGAACCACCACAGCCATCAACATCGACGTCAGAACCTCTTCCCACAGCCTCGGCC 666

Qy    661 TCTTCCACGGAGCCTTCTCCTGCGAGGGCGAGAGCCGTTCCTCCAGTGTTGTGGGTCTGGGGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    667 TCTTCCACGGAGCCTTCTCCTGCGAGGGCGAGAGCCGTTCCTCCAGTGTTGTGGGTCTGGGGGT 726

Qy    721 GGTGGCATCCCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    727 GGTGGCATCTCCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTT 786

Qy    781 GCCTCCAGCTCTCCAGACAGAAAGACTGGGTCCTTTTCGTTGGAACCCCAGGATCAG 840
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    787 GCCTCCAGCTCTCCAGACAGAAAGACTGCGTCCTTTTCGTTGGAACCCCAGGATCAG 846

Qy    841 GAGGATTTGGAGCCCGTGAACAAAATGAGAGGAGATGGGGACCTTGACCTTGAACGG 900
              |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db    847 GAGGATTTGGAGCCCGTGAACAAAATGAGAGGAGATGGGGA-CTTGACCTTGAACGG 905

Qy    901 CAGTTGTTGGTCGCACAACCCGTAGAAATGCCCAAACCGTCCACGAGGACGTCAGAGCA 960
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    906 CAGTTGTTGGTCGCACAACCCGTAGAAATGCCCAAACCGTCCACGAGGACGTCAGAGCA 965

Qy    961 GCGGCTGGGAAGCCAGACAAGATGGAGGACGCTGACATCACTCTGCCACGACCTG 1020
```

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2559 row: o   column: 18
High quality sequence stop: 678.

FEATURES   Location/Qualifiers
source   1..876
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6376937"
                /tissue_type="normal pigmented retinal epithelium"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC 43"
                /note="Organ: eye; Vector: pOTB7; Site 1: XhoI; Site 2:
                EcoRI; cDNA made by oligo-dT priming. Directionally
                cloned into EcoRI/XhoI sites using the following 5'
                adaptor: GGCACGAG(G). Library constructed by Ling Hong
                in the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library. |"

ORIGIN

Query Match   30.7%;   Score 822.2; DB 13; Length 876;
Best Local Similarity 98.3%;   Pred. No. 1.4e-126;
Matches 863; Conservative   0; Mismatches   0; Indels   6; Gaps   3;

```
Qy  1165 GTGGAAGCATACCTCATCCAGCATCCAGACAGAGTCGCAGTGAAGAGATGTGCAAAGT 1224
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GTGGAAGCATACCTCATCCAGCATCCAGACAGAGTCGCAGTGAAGAGATGTGCAAAGT 60

Qy  1225 ATGATGCCAGGAATAAAATCACTTCAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTT 1284
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ATGATGCCAGGAATAAAATCACTTCAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTT 120

Qy  1285 TCTGATGAAGAAGGGAGTTCAGGAGACCTTGTCAGAGCGTTGACAGTGAGTCC 1344
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TCTGATGAAGAAGGGAGTTCAGGAGACCTTGTCAGAGCGTTGACAGTGAGTCC 180

Qy  1345 TCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCG 1404
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCG 240

Qy  1405 GCGCAGCCTCCCCACTGCCCAGGAGCACCCGAGGGGGAGCACCCAGGCCCTGGGG 1464
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GCGCAGCCTCCCCACTGCCCAGGAGCACCCGAGGGGGAGCACCCAGGCCCTGGGG 300

Qy  1465 GATGCACCCTCCAGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTGTGCCCCTCTG 1524
         ||||||||||||||||||||||||||||||||||||||||   :|||||||||||||||
Db   301 GATGCACCCTCCAGTCCGTCAGCCTGACGACAGCA---CAGTCCAGGATTACGTGTGCCCCTCTG 357

Qy  1525 CAAGGAAGCCACGGCCCTGCACCTGCTTCCAGCCATGCCCGACGGAGAGCGGAG 1584
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   358 CAAGGAAGCCACGGCCCTGCACCTGCTTCCAGCCATGCCCGACGGAGAGCGGAG 417

Qy  1585 CGCGGCAGGACACCCGCGTGCGCCCTCAGCAGTGTGCGGTCTGCAGCCTTTCTGC 1644
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   418 CGCGGCAGGACACCCGCGTGCGCCCTCAGCAGTGTGCGGTCTGCAGCCTTTCTGC 477

Qy  1645 CACCTGTACTGGGGCTGACACGGCTGCCTGGCCCCGTTTTGTGAG 1704
         ||||||||||||||||||||||||||||||||||||||||||||
Db   478 CACCTGTACTGGGGCTGCACCGGGACCGGCTGCCTGGCCCCGTTTTGTGAG 537

Qy  1705 CTCAACCTGGGTGACAAGTGTCTGGACAGGATCTGAACAACACAGCTACGAGTCAGAC 1764
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   538 CTCAACCTGGGTGACAAGTGTCTGGACAGGATCTGAACAACACAGCTACGAGTCAGAC 597

Qy  1820 AGAGCCCTCGTGGCCTCCAGCGGGGAGTGTTTCTGCGTCTGATTACAGAGTCACGGGAG 1879
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   824 AGAGCCTCGTGGCCTCCAGCGGGGAGTGTTTCTGCGTCTGATTACAGAGTCACGGGAG 765

Qy  1880 ACACCGTTCTGTGTTACTGCTGTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAGTATC 1939
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   764 ACACCGTTCTGTGTTACTGCTGTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAGTATC 705

Qy  1940 GGCAGAAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCCGTCCTGACTGCTACT 1999
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   704 GGCAGAAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCCGTCCTGACTGCTACT 645

Qy  2000 GGGGCCGTAACTGCCGCACTCAGGTGAAAGCTCACCACCGCCATGAAATTCAATCATATCT 2059
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   644 GGGGCCGTAACTGCCGCACTCAGGTGAAAGCTCACCACCGCCATGAAATTCAATCATATCT 585

Qy  2060 GTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTCAGCACTGG 2119
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   584 GTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTCAGCACTGG 525

Qy  2120 AGGTGAAGAGAGCGTGTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTTCACAGCC 2179
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   524 AGGTGAAGAGAGCGTGTTTTTAAAATACAGAGACACAGCACGTCAAGGTGTTTTCACAGCC 465

Qy  2180 CCCTGAGGGAAGGGACGCAGGGTCTCCGACAGGTGCTCTGGGTGACTCTTCGTGGAGC 2239
         ||:||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   464 CCCTGAGGGAAGGGACGCAGGGTCTCCGACAGGTGCTCTGGGTGACTCTTCGTGGAGC 405

Qy  2240 TTTTTACCCCTGAGTGAGACCCTCCCAGAGCCCCGGGGCCCGGGGGCCCGAGCCCCTCCTGG 2299
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   404 --TTTACCCCTGAGTGAGACCCTCCCAGAGCCCCGGGGCCCGGGGGCCCGAGCCCCTCCTGG 347

Qy  2300 TGAGCGGCTGGGCAGGGCTCGTGGTGGCATCAGCAGCAGCAGACGAAGCCTTTCTGTAACAT 2359
         ||||||||||||:|:||||||||||||||||||||||:|:|||||||||||:|||||||||
Db   346 TGAGCGGCTGSGCAGVGCCCGGTGGCTGCMTCAGCMGCAGAGACGAASCCYYTCTGTMACAT 287

Qy  2360 GCGGGCCGTCCGCGCCAGAGAGGGGCAGTTTTGCTCTTTTGTACATTTCGAAACTACAGTT 2419
         ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   286 GCGSCCSTCCGCGCCAGAGAGGGGGMARTTTYGCTCTTTTGTACATTTCGAAACTACAGTT 227

Qy  2420 AAAGCAGAAGTCTGTTTTCAGGAAAAGTTTCAAGGGAGAAGGGCAAGTTTATCAAAAACA 2479
         ||||||||||:|||||||||||||||||:::|||||||||||||||||::||||||||
Db   226 AAAGCCGAAGTCTGTTTTCAGGAAWWKTTTCAAGGGAGAAGGGCAAKTTTTTTTAAWACA 167

Qy  2480 TTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTACAGGACGTACACAATATCCTGCTG 2539
         |||||||||||||||||||||||||||:|::|||||||||||||:|||||||||||:||
Db   166 TTGTTTCAGGAGAAGGGKGCATWDGTTTACWGCCTACAGGACGTABACAATATCCTGCTG 107

Qy  2540 CTGGGAAAACACCACACATTTTATCTATTTTTATTTTAAATAGGTTGGTGCTTATCTTCT 2599
         |:::|||||   ::||||:||||||:|||||||||||:||||:||||||::|||||:|||
Db   106 CTKKGAAAGGSSACGGGGTTTTTATSTDTTTTTATTTTAAATAGGTTTGGTGCTTKTCTTTCT 47

Qy  2600 AATAAGATTTAAAATGTCACAAAACTGTAGCACAAATAAATAAATTTAT 2646
         |||||||||||||||||     :|:||:||||||||||||||||||
Db    46 AATAAGATTTAAAATGTCACATAKTGTA-YACWAWTAATATAATTTAW 1
```

RESULT 5
BQ876252
LOCUS   BQ876252   876 bp   mRNA   linear   EST 16-AUG-2002
DEFINITION AGENCOURT_8681666 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:6376937
           5', mRNA sequence.
ACCESSION BQ876252
VERSION   BQ876252.1 GI:22268258
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 876)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)

```
Qy  1765 ATCCTGAAGAATTACCTGGCAACCAGAGGTTTGACATGGAAAAACATGTTGACCGAGAGC 1824
         |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   598 ATCCTGAAGAATTACCTGCAACCAGAGGTTTGACATGGAAAAACATGTTGACCGAGAGC 657

Qy  1825 CTCGTGGCCTCCAGCGGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGG--AGACA 1882
         ||| |||| ||||||||||||||||||||||||||||||||||||||||||||  ||||
Db   658 CTCATGGCTCTCCAGCGGGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACAC 717

Qy  1883 CCGTTCTGTGTTACTGCTGTGTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGC 1942
         |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   718 CCGTTCTGTTACTGCTGTGTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGC 777

Qy  1943 AGAACATTCCTGCTTCCGAGTGCCAGTGCCGTAACATCCGTCCGACTGCTACTGGG 2002
         |||||||||||||||||||| |||||| |||||||||| |||||||||||||||||
Db   778 AGAACATTCCTGCTTCCGAGNTGCCAGTGGCCGTAACATCCCGTCCGACTGCTACTGGG 837

Qy  2003 -GCCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGC 2039
          ||||||||||||||| ||||||||||||||||||||
Db   838 NGGCGTAACTGCCGCACTCCGGTGAAAGCTCACCACGC 875
```

RESULT 6
BU176559
LOCUS   BU176559   921 bp   mRNA   linear   EST 04-SEP-2002
DEFINITION   AGENCOURT_7940317 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6143609
              5', mRNA sequence.
ACCESSION   BU176559
VERSION   BU176559.1   GI:22690543
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 921)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM13467 row: e column: 18
              High quality sequence stop: 697.
FEATURES   Location/Qualifiers
source   1..921
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:6143609"
              /tissue_type="retinoblastoma"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_67"
              /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
              Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
              Average insert size 1.75 kb. Library constructed by Life
              Technologies."
ORIGIN

Query Match   28.9%;   Score 774.8;   DB 13;   Length 921;
Best Local Similarity 99.7%;   Pred. No. 1e-118;
Matches 776;   Conservative   0;   Mismatches   2;   Indels   0;   Gaps   0;

```
Qy  1889 TGTGTTACTGCTGTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACA 1948
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 TGTGTTACTGCTGTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACA 60

Qy  1949 TTCCTGCTTCCGAGTTGCCAGTGCCGTAACATCCGTCCGACTGCTACTGCGGCCGTA 2008
```

```
Db    61 TTCCTGCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCGACTGCTACTGGGGGCCGTA 120
         |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Qy  2009 ACTGCCGCACTCAGGTGAAAGCTCACCACGCACCACGCCATGAAATTCAATCATATCTGTGACACAGA 2068

Db   121 ACTGCCGCACTCAGGTGAAAGCTCACCACGCACCACGCCATGAAATTCAATCATATCTGTGACACAGA 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2069 CAAGGTTCAAAAACTAAGCATCCAGAGCCCTGAGCAGCTTTCAGCACTGGAGGTGAAGA 2128

Db   181 CAAGGTTCAAAAACTAAGCATCCAGAGCCCTGAGCAGCTTTCAGCACTGGAGGTGAAGA 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2129 GAGCGTGTTTTAAAATACAGAGACAGCAGCACGCACGCAAGGTGTTTCACAGCCCCCTGAGGG 2188

Db   241 GAGCGTGTTTTAAAATACAGAGACAGCAGCACGCACGCAAGGTGTTTCACAGCCCCCTGAGGG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2189 AAGGGACGGACGGGTCTCCGACAGGTGCTCCTGGGGTGACTCTTCTGTGGGAGCTTTTACCC 2248

Db   301 AAGGGACGGACGGGTCTCCGACAGGTGCTCCTGGGGTGACTCTTCTGTGGGAGCTTTTACCC 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2249 TCTGAGTGAGACCCTCCCCAGAGCCCCGGGGCCCGGGGGCCCGCAGCCCGCCCTCCTGGTGAGCGCTG 2308

Db   361 TCTGAGTGAGACCCTCCCCAGAGCCCCGGGGCCCGGGGGCCCGCAGCCCGCCCTCCTGGTGAGCGCTG 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2309 GGCAGGGCTCGTGTGGCATCAGCAGCGAGCAGGAGACGAAGCCTTTCTGTAACATGCGGCCGTC 2368

Db   421 GGCAGGGCTCATGTGGCATCAGCAGCAGCAGAGACGAAGCCTTTCTGTAACACATGCGGCCGTC 480
         |||||||||| ||||||||||||||| || |||||||||||||||||||||| |||||||||||
Qy  2369 CGCCGAGAGGGGCAGTTTGCTTTTGTACATTTTGTACATTTTCCGAAACTACAGTAAAGCAGAA 2428

Db   481 CGCCGAGAGGGGCAGTTTGCTTTTGTACATTTTGTACATTTTCCGAAACTACAGTAAAGCAGAA 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2429 GTCTGTTTCAGGAAAAGTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAG 2488

Db   541 GTCTGTTTCAGGAAAAGTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2489 GAGAAGGGAGCATAAGTTTACAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAA 2548

Db   601 GAGAAGGGAGCATAAGTTTACAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAA 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2549 CCACAGCATTTATCTATTTTTTATTTTAATAGGTTGGTGTGCTTATCTTCTAATAAGATT 2608

Db   661 CCACAGCATTTATCTATTTTTTATTTTAATAGGTTGGTGTGCTTATCTTCTAATAAGATT 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2609 TAAATGTCACAAACTGTAGCACAAATAATAATATAATTTATAATTTACAAATTGACTAAAA 2666

Db   721 TAAATGTCACAAACTGTAGCACAAATAATATAATTTATAATTTACAAATTGACTAAAA 778
```

RESULT 7
BQ215731
LOCUS   BQ215731   870 bp   mRNA   linear   EST 02-MAY-2002
DEFINITION   AGENCOURT_7549119 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6059147
              5', mRNA sequence.
ACCESSION   BQ215731
VERSION   BQ215731.1   GI:20397131
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 870)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: DCTD/DTP/Gazdar
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov

```
Qy   1695  GTTTTGTGAGCTCAACCTGGGTG 1717
            |||||||||||||||| ||||
Db    778  GTTTTGTGAGCTCCACCTGGGG 800
```

RESULT 8
BQ073354
LOCUS   BQ073354   992 bp   mRNA   linear   EST 02-APR-2002
DEFINITION AGENCOURT_7052199 NIH_MGC_101 Homo sapiens cDNA clone IMAGE:5806143
            5', mRNA sequence.
ACCESSION   BQ073354
VERSION   BQ073354.1 GI:19902400
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 992)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2046 row: h column: 16
            High quality sequence stop: 670.

FEATURES   Location/Qualifiers
     source   1..992
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5806143"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_101"
                     /note="Organ: lung; Vector: pOTB7; Site_1: EcoRI; Site_2:
                     XhoI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."

ORIGIN

Query Match   28.3%;   Score 758; DB 13; Length 992;
Best Local Similarity 96.6%; Pred. No. 6.3e-116;
Matches 817; Conservative 0; Mismatches 25; Indels 4; Gaps 4;

```
Qy  551  GGGGAGGGGCGATCCCCGGGTCCCCTCCGGTGTCGCGCCCCCACTCAGGTGTGCTTTGAGG 610
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    1  GGGGAGGGGCGATCCCCGGGTCCCCTCCGGTGTCGCGCCCCCACTCAGGTGTGCTTTGAGG 60

Qy  611  AACCACAGCCATCAACATGACGTCAGACCTCTTCCCACAGCCTCGGCCTCTTCCACGG 670
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  AACCACAGCCATCAACATGACGTCAGACCTCTTCCCACAGCCTCGGCCTCTTCCACGG 120

Qy  671  AGCCTTCTCCTGCAGGGCGGAGAGCGTTCCCAGTGTGGGTCTGGGGTGGTGGCATCT 730
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  AGCCTTCTCCTGCAGGGCGGAGAGCGTTCCCAGTGTGGGTCTGGGGTGGTGGCATCT 180

Qy  731  CCCCTAAAGGAAGTGGTCCCTCGTGTGGCAAGTGATGAAGTCTCCAAGTCCTCAGCTC 790
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  CCCCTAAAGGAAGTGGTCCCTCGTGTGGCAAGTGATGAAGTCTCCAAGTCCTCAGCTC 240

Qy  791  TCCCAGACAGAAAAGACTGGGTCCTTTTCGTTGGAACCCCAGGATCAGGAGGATTTGG 850
          |||||||||||||||||||||||||| |||||||||||| ||||||||||||| |||||
Db  718  CCTTTCTGCCACCTGTACTGGGGCTGCTACGGCCGGCTGCTACCTGTGCTGGNCCCCC 777
```

Plate: LLAM13325 row: f column: 12
High quality sequence stop: 601.
FEATURES   Location/Qualifiers
     source   1..870
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6059147"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_68"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.8 kb. Library constructed by Life
                     Technologies."

ORIGIN

Query Match   28.6%;   Score 765.4; DB 13; Length 870;
Best Local Similarity 98.8%; Pred. No. 3.9e-117;
Matches 793; Conservative 0; Mismatches 6; Indels 4; Gaps 2;

```
Qy  916  CAACCGCTGAGAAATGCCCAAACCGTCCACGAGGACGTCAGAGCAGCAGCGGCTGGGAAGCCA 975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  CAACCGCTAGAAATGCCCAAACCGTCCACGAGGACGTCAGAGCAGCAGCGGCTGGGAAGCCA 60

Qy  976  GACAAGATGGAGGAGACGCTGACATGCATCATCTGCCAGGACCTGCTGCACGACTGCGTG 1035
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  GACAAGATGGAGGAGACGCTGACATGCATCATCTGCCAGGACCTGCTGCACGACTGCGTG 120

Qy 1036  AGTTTGCAGCCCTGCATGCACACGTTCTGCGGCTTGCTACTCGGCTGGATGGAGCGC 1095
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  AGTTTGCAGCCCTGCATGCACACGTTCTGCGGCTTGCTACTCGGCTGGATGGAGCGC 180

Qy 1096  TCGTCCCTGTGTCTCCTACCCTGCCGCGTGTCCCGTGAGCGGATCTGTAAAAACCACATCCTC 1155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TCGTCCCTGTGTCCTCACCTGCCGCGTGTCCCGTGAGCGGATCTGTAAAAACCACATCCTC 240

Qy 1156  AACAACCTGTGGAAGCATACCTCCATCCAGCATCCAGCAAGAGTCGCAGTGAAGAAGAT 1215
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  AACAACCTGTGGAAGCATACCTCCATCCAGCATCCAGCAAGAGTCGCAGTGAAGAAGAT 300

Qy 1216  GTGCAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGTCCAGCCCAAGTCAGG 1275
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GTGCAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGTCCAGCCCAAGTCAGG 360

Qy 1276  CGGTCTTTTCTGATGAAGAAGGAGTTCAGAGAACCTGCTGGAGCTGTCAGACGTTGAC 1335
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  CGGTCTTTTCTGATGAAGAAGGAGTTCAGAGAACCTGCTGGAGCTGTCAGACGTTGAC 420

Qy 1336  AGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTCGCCGGCAGTGTCCTGAGTACAGA 1395
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  AGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTCGCCGGCAGTGTCCTGAGTACAGA 480

Qy 1396  AGGCAGGCGGCGCAGCCTCCCACTGCCCAGCACCCGAGGGGCGAGCCAGGAGCCCACAG 1455
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  AGGCAGGCGGCGCAGCCTCCCACTGCCCAGCACCCGAGGGGCGAGCCAGGAGCCCACAG 540

Qy 1456  GCCCTGGGGATGCACCCTCAGCGTCCGTCAGCAGCAGCAGGATTACGTG 1515
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  GCCCTGGGGATGCACCCTCAGCGTCCGTCAGCAGGA---CAGTCCAGGATTACGTG 597

Qy 1516  TGCCCTCGCAAGGAAGCCACGCCGTGCACCTGCCTCAGCCATGCCCGACCGG 1575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  598  TGCCCCTCGCAAGGAAGCCACGCCGTGCACCTGCCTCAGCCATGCCCGAACGG 657

Qy 1576  AGAGCGGAGCGCGAGCAGGAGACCCGGCTGCGGTGTGCCGGAG 1635
          |||||||||||||||||||||||||||||||||||||||||||
Db  658  AGAGTGAGCGGAGCGCGAGCAGGAGACCCGGCTGCGGTCGGCAG 717

Qy 1636  CCTTTTCTGCCACCTGTACTGGGGCTGCTACGGCCGGCTGCTA-CCCC 1694
          |||||||||||||||||||||||||||||||||||||||||||||
Db  718  CCTTTTCTGCCACCTGTACTGGGGCTGCTACGGCCGGCTGCTAGGNCCCCC 777
```

```
Db    241  TCCCAGACAGAAAGACTCGTCCTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTTGG  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    851  AGCCCGTGAAGAAGAAAATGAGAGGAGAATGGGAGACCTTGACCTGAACGGGCAGTTGTTGG  910
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  AGCCCGTGAAGAAGAAAATGAGAGAGAGATGGGAGACCTTGACCTGAACGGGCAGTTGTTGG  360

Qy    911  TCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAGGACGTCAGAGCACGCGGCTGGGA  970
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  TCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAGGACGTCAGAGCACGCGGCTGGGA  420

Qy    971  AGCCAGACAAGATGGAGGAGACCGTGACATGCATCTGCCAGGACCTGCTCGCACGACT  1030
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  AGCCAGACAAGATGGAGGAGACCGTGACATGCATCATCTGCCAGGACCTGCTCGCACGACT  480

Qy   1031  GCGTGAGTTGCAGCCCTGCATGCACACGTTCTGCCGCGGCTACTCGGCGTGGATGG  1090
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  GCGTGAGTTGCAGCCCTGCATGCACACGTTCTGCCGCGGCTACTCGGCGTGGATGG  540

Qy   1091  AGGCCTCGTCCCCTGTCCTACCTGCGCGTCCCGTGGAGCGGATCGTGTAAAAACCACA  1150
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  AGGCCTCGTCCGTCCCTGTGTCCTACCTGCGCGTCCCGTGGAGCGGATCGTGTAAAAACCACA  600

Qy   1151  TCCTCAACAACCTCGTGGAAGCATACCTTCATCCAGCATCCAGACAAGAGTCGCAGTGAAG  1210
           |||||||||||||||||||||||||||||||||||||||||||||
Db    601  TCCTCAACAACCTCGTGGAAGCATACCTTCATCCAGCATCCAGACAAGAGTCGCAGTGAAG  660

Qy   1211  AAGATGTGCAAAGTAATGATGATGCCAGGAATAAAAATCACTCAAG-ACATGCTGCAGCCCAAA  1269
           ||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    661  AAGATGTGCAAAGTAATGATGATGCCAGGAATAAAAATCACTCAAGACACATGCTGCAGCCCAAA  720

Qy   1270  GTCAGGCGCGTCTTTTCTGATGAAGAAGAGGAGTTCAGAGAACCTGCT-GGAGCTGTCAG-  1327
           ||||||||||||||||||||||||||||||||||||||| ||||||||| |
Db    721  GTCAGGCGCGTCTTTTTCTGATGAAGAAGAGGAGTTCAGAGAACCTGCTGGGGACCTGTCAGA  780

Qy   1328  ACGTTGACAGTGAG-TCCTCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCT  1386
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    781  ACGTTGACAGTGAGTTCCTCAAACATTAAAGCCAGCCCTACCTTCTTGGGGCCGGGGCAGT  840

Qy   1387  GAGTAC  1392
           ||||
Db    841  GGGTCC  846


RESULT 9
CD243458
LOCUS       CD243458                877 bp    mRNA         linear    EST  22-MAY-2003
DEFINITION  AGENCOURT 14121909 NIH_MGC 180 Homo sapiens cDNA clone
            IMAGE:30383176 5', mRNA sequence.
ACCESSION   CD243458
VERSION     CD243458.1   GI:31003922
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 877)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM446 row: j  column: 17
            High quality sequence stop: 660.
FEATURES             Location/Qualifiers
     source          1..877
                     /organism="Homo sapiens"
```

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30383176"
                     /lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC 180"
                     /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site 1: NotI;
                     Site 2: EcoRV (destroyed) Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.68 kb. Library was
                     constructed by (Invitrogen). Note: this is a NIH_MGC
                     Library."

ORIGIN

Query Match    28.3%;    Score 757.6;  DB 14;   Length 877;
Best Local Similarity  96.9%;    Pred. No. 7.6e-116;
Matches  772;  Conservative   0;  Mismatches  25;  Indels  0;  Gaps  0;

Qy   1527  AGGAAGCCACGCCCTGTGCACCTGCTTCCAGCCCATGCCGACCCGAGAGCCGGAGCGGAGCG  1586
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4  AGGAAGCCACGCCCTGTGCACCTGCTTCCAGCCCATGCCGACCCGACCGGAGAGCCGGAGCGGAGCG  63

Qy   1587  CGAGCAGGAACCCGCGTGTCGCCCCTCAGCGAGTGTGGGCGTCGCCTGGGCCCGTTTCTGCCA  1646
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     64  CGAGCAGGAACCCGCGTGTCGCCCCTCAGCAGTGCGGGTGTGCCTGCCTTCTGCCA  123

Qy   1647  CCTGTACTGGGGTGACCCGGCTGCTCTGAACGGCGTGCCTGGCCCCGTTTTGTGTGAGCT  1706
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    124  CCTGTACTGGGGTGACCCGGCTGCTCTGAACGGCGTGCCTGGCCCCGTTTTGTGTGAGCT  183

Qy   1707  CAACCTGGGTGACAAGTGTCTGGACACGGCGTGCTGAACAACAACAGCTACGAGTCAGACAT  1766
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184  CAACCTGGGTGACAAGTGTCTGGACACGGCGTGCTGAACAACAACAGCTACGAGTCAGACAT  243

Qy   1767  CCTGAAGAATTACCTGGCAACCAGAGTTTGACATGGAAAAACATGTTGACGGGAGCACCGT  1826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    244  CCTGAAGAATTACCTGGCAACCAGAGTTTGACATGGAAAAACATGTTGACGGGAGCACCGT  303

Qy   1827  CGTGGCTTCCAGCGGGGAGTGTTTCTGCGTGCTTCCGTGAGCTTCAGCTCATCAGTATCGGCAGAA  1886
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304  CGTGGCTTCCAGCGGGGAGTGTTTCTGCGTGCTTCCGTGAGCTTCAGCTCATCAGTATCGGCAGAA  363

Qy   1887  TCTGTGTTACTGCTGTGGCCTGCGCAGCTTCCGCGAGCTGCCAGTGGCCGTAACATCCCGTCCTGACTGCTCTACTGGGGCG  1946
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    364  TCTGTGTTACTGCTGTGGCCTGCGCAGCTTCCGCGAGCTGCCAGTGGCCGTAACATCCCGTCCTGACTGCTCTACTGGGGCG  423

Qy   1947  CATTCCTGCTTCCGAGTTGCCAGTGGCCGTGGCCGTAACATCCCGTCCTGACTGCTCTGACTGCTCTACTGGGGCCG  2006
           |||||||||||||||||||||||||||||||||||||||||||||
Db    424  CATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCCGTCCTGACTGCTCTACTGGGGCCG  483

Qy   2007  TAACTGCCGACTCAGGTGAGAAGCTCACCACGCCATGAAATTCAATCATATCTGTGAACA  2066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    484  TAACTGCCGACTCAGGTGAGAAGCTCACCACGCCATGAAATTCAATCATATCTGTGAACA  543

Qy   2067  GACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAGGTGAA  2126
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    544  GACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAGGTGAA  603

Qy   2127  GAGAGGCGTGTTTAAATACAGAGACAAGACACGTCAAGGTGTTTCACAGCCCCTGAG  2186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    604  GAGAGGCGTGTTTAAATACAGAGACAAGACACGTCAAGGTGTTTTCACAGCCCCTGAG  663

Qy   2187  GGAAGGGACGGACGGCAGGTGCTCTCGGGGTGACTCTTCTGTGGAGCTTTTAC  2246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    664  GGAAGGGACGGACGGCAGGTGCTCTCGGGGTGACTCTTCTGTGGAGCTTTTTAC  723

Qy   2247  CCTCGAGTGAGACCCTCCCAGAGCCCGGGGGCCGGACCCCCTCGGTGAGCGC  2306
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    724  CCTCGAGTGAGACCCTCCCAGAGCCCGGGGGCCGCCAGCCCCAGCCCCTCCGTGTGA  783

Qy   2307  TGGGCAGGGCTCGTGT  2323
           |||| ||||
Db    784  GCGCTTGGGCAAGGGCT  800
```

RESULT 10
BU148471
LOCUS   BU148471   910 bp   mRNA   linear   EST 03-SEP-2002
DEFINITION AGENCOURT_8675647 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6380102
            5', mRNA sequence.
ACCESSION   BU148471
VERSION   BU148471.1 GI:22662003
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 910)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2568 row: c column: 15
            High quality sequence stop: 620.
FEATURES   Location/Qualifiers
     source   1..910
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6380102"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_40"
                     /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

ORIGIN

Query Match   27.6%;   Score 740.6; DB 13; Length 910;
Best Local Similarity 91.1%;   Pred. No. 5e-113;
Matches 858; Conservative 0; Mismatches 40; Indels 44; Gaps 5;

Qy   12 ACGAGGCCGCAATGTCTCTTGACACAGCGGCGGCGGCGCCGGTTCCGGGTTCGGCGGCGG 71
          | |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   5 AGGGAGCCGCAATGTCTCTTGACACAGCGGCGGCGGCGCCGGTTCCGGGTTCGGCGGCGG 64

Qy   72 GGCGGGGGATGTGAATCCCGATGGAGCGGGCCCCGAGGAAGGCAAGCAGTCGCCGCGCCGCCCA 131
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 GGCGGGGGATGTGAATCCCGATGGAGCGGGCCCCGAGGAAGGCAAGCAGTCGCCGCGCCGCCCA 124

Qy   132 GCCCTGGGGACGGCTCCTGCGTCTGGGCGCGGAGGAGGGGCAGCCGCCAGTCCTCCTGAG 191
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   125 GCCCTGGGGACGGCTCCTGCGTCTGGGCGCGGAGGAGGGGCAGCCGCCAGTCCTCCTGAG 184

Qy   192 GAAGGCGGAGTGGACCATCGGGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAA 251
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   185 GAAGGCGGAGTGGACCATCGGGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAA 244

Qy   252 ACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAGAAAATCAGGTGACGGTGACACT 311
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   245 ACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAGAAAATCAGGTGACGGTGACACT 304

Qy   312 GGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGAC 371
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   305 GGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGAAGCAGAC 364

Qy   372 ATGCCCTTTACAGACTGGGGATGTCATCTACTTGTGTACAGGAAGAATGAACCGGAACA 431
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   365 ATGCCCTTTACAGACTGGGGATGTCATCTACTTGTGTACAGGAAGAATGAACCGGAACA 424

Qy   432 CAACGTGGCATACCTCTATGAATCTTAAGTGAAAAGCAAGGCATGACACAAGAATCCTT 491
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   425 CAACGTGGCATACCTCTATGAATCTTAAGTGAAAAGCAAGGCATGACACAAGAATCCTT 484

Qy   492 TGAAGCTAACAAGGAAAAATGTGTTCCATGGACCAAAGATAACTCAGGTGCAGGTGCAGG 551
          ||||||||||||||||||| |||||||||||||||||||
Db   485 T------------------------------------GATACCTCAGGTGCAGGTGCAGG 508

Qy   552 GCGAGGGGGCCGATCCCCGGTCCCTCCGTCGTCGGCCGCCCACTCAGGTGTGCTTTGAGGA 611
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   509 GCGAGGGGGCCGATCCCCGGTCCCTCCGTCGTCGGCCGCCCACTCAGGTGTGCTTTGAGGA 568

Qy   612 ACCACAGCCATCAACATCGACGTCAGACCTCTTCCCACAGCCTCGGCCTCTTCCCACCGA 671
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   569 ACCACAGCCATCAACATCGACGTCAGACCTCTTCCCACAGCCTCGGCCTCTTCCCACCGA 628

Qy   672 GCCTTCCTGCAGGGGCAGGAGAGCGGTTCCTCCCAGTTGTGGGTTCTGGGGGTGGTGGCATCTC 731
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   629 GCCTTCCTGCAGGGGCAGGAGAGCGGTTCCTCCCAGTTGTGGGTTCTGGGGGTGGTGGCATCTC 688

Qy   732 CCCTAAAGGAAGTGGTGTCCCTCGTGGCAAGTGATGATGAAGTGCTCCAGCTCT 791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   689 CCCTAAAGGAAGTGGTGTCCCTCGTGTGGCAAGTGATGATGAAGTGCTCCAGCTCT 748

Qy   792 CCAGACAGAAAGACT-GGTCCTTTTCGTCGTTGGAACCCCAGGATCAGAGGAGGATTGG 850
          |||| |||||||||| |||||||||| |||||||| ||||||||||||||||||||||
Db   749 CCCAAACAGAAAGACTGGGTCCTTTTCGTCGGTGGAACCCCAGGATCAGGGGGGGATTTG 808

Qy   851 AGCCCGTGAAGAAGAAAATGA----GAGGAGATGGGGACCTTG--ACCTGAACGGGCAGT 904
          |||||||||||||||||||| |||||||||||| |||||| |||||||||||||
Db   809 GGAGCCCGTGAAGAAGAAAATGACAACAAGGAGATGGGGACCCTGTACCTGAACGGGGCAGT 868

Qy   905 TGTTGGTCGCACAACC-CGGTAGAAATGCCAAACCGTCCAC 945
          |||||||||||||||| ||||||||||||||||||||||
Db   869 GGTGGTTCCCACAACCACCGGGGGTAGAAATGCCCCAAACGGGGCCC 910

RESULT 11
BG395714
LOCUS   BG395714   840 bp   mRNA   linear   EST 12-MAR-2001
DEFINITION 602458365F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4580633 5',
            mRNA sequence.
ACCESSION   BG395714
VERSION   BG395714.1 GI:13289162
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 840)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1300 row: a column: 18
            High quality sequence stop: 775.
FEATURES   Location/Qualifiers
     source   1..840
                     /organism="Homo sapiens"
                     /mol_type="mRNA"

Db   782 AGGTTCAAAACTCAGGCATCCAGAGGCCTGAAG 814

RESULT 12
BQ936072
LOCUS   BQ936072
DEFINITION   AGENCOURT 8779684 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6371968
             5' mRNA sequence.
ACCESSION   BQ936072
VERSION   BQ936072.1 GI:22351455
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 953)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM2546 row: p column: 17
             High quality sequence stop: 630.
FEATURES   Location/Qualifiers
     source   1..953
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6371968"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_18"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."

ORIGIN

Query Match   27.1%;   Score 726.8; DB 13; Length 953;
Best Local Similarity 93.0%; Pred. No. 9.6e-111;
Matches 842; Conservative 0; Mismatches 18; Indels 45; Gaps 6;

Qy   45 GCGCAGCCGGTTCCGGTTCGGCGCCGCGCCGCGGGGGGATGTGAATCCGATGGAGCGGCCCGA 104
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGGCAGCCGGTTCCGGTTCGGCGCCGCGCCGCGGGGGGATGTGAATCCGATGGAGCGGCCCGA 60

Qy   105 GGAAGGCAAGCAGTCGCCGCGCCGCGCCGCGCCCTGGGACGGGCTCTCGGGCGCGCGGA 164
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGAAGGCAAGCAGTCGCCGCGCCGCGCCGCGCCCTGGGACGGGCTCTCGGGCGCGCGGA 120

Qy   165 GGAGGGCGAGCCGCACGTCCTCCTGAGGAAGCGGAGTGGACCATCGGGCGGAGACGAGG 224
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGAGGGCGAGCCGCACGTCCTCCTGAGGAAGCGGAGTGGACCATCGGGCGGAGACGAGG 180

Qy   225 TTGCGACCTTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGATCACTCGTAGAATTGTAGT 284
         |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   181 TTGCGACCTTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGATCACTCGTAGAATTGTAGT 240

Qy   285 GGATGAAAAATCAGGTCAGGTGACACTGGAAGATACCAGCACCAGTCGAACAGTGATTAA 344
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GGATGAAAAATCAGGTCAGGTGACACTGGAAGATACCAGCACCAGTCGAACAGTGATTAA 300

                     /db_xref="taxon:9606"
                     /clone="IMAGE:4580633"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_16"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

ORIGIN

Query Match   27.2%;   Score 727.4; DB 12; Length 840;
Best Local Similarity 95.8%; Pred. No. 8e-111;
Matches 779; Conservative 0; Mismatches 31; Indels 3; Gaps 3;

Qy 1294 GAAGGGAGTTCAGAGACCTGCTGGGAGCGTGTCAGACGTTGACAGTGAGTCCTCAGACATT 1353
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 GAAGGGAGTTCAGAGACCTGCTGGGAGCGTGTCAGACGTTGACAGTGAGTCCTCAGACATT 61

Qy 1354 AGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAAGGCGGGCGCGCAGCCT 1413
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 AGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAAGGCGGGCGCGCAGCCT 121

Qy 1414 CCCCACTGCCCAGCACCCGAGGGCGAGCCAGGAGAGCCCCACAGGCCCTGGGGATGCACCC 1473
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 CCCCACTGCCCAGCACCCGAGGGCGAGCCAGGAGAGCCCCACAGGCCCTGGGGATGCACCC 181

Qy 1474 TCCACGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTGTGCCCCTCTGCAAGGAAGC 1533
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 TCCACGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTGTGCCCCTCTGCAAGGAAGC 241

Qy 1534 CACGCCCTGTGCACCTGCTGCTTCCAGCCCATGCCCGACCGGAGACGGCGGAGCAG 1593
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 CACGCCCTGTGCACCTGCTGCTTCCAGCCCATGCCCGACCGGAGAGTGGAGCGCGAGCAG 301

Qy 1594 GACCCGGCGTGTCGCCCCTCAGCAGTGTGCGGTCTGCCTGCCGCAGCCTTTCTGCCACCTGTAC 1653
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 GACCCGGCGTGTCGCCCCTCAGCAGTGTGCGGTCTGCCTGCCGCAGCCTTTCTGCCACCTGTAC 361

Qy 1654 TGGGGCTGCACCCGAGGAGGGTGCACCCGGTGGGCTGCACCTGGAGGTGAGAGTCACCTG 1713
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362 TGGGGCTGCACCCGAGGACCGGGTGCTACGGCTGCTACGGCTGCTGGAGTGAGAGTCACCTG 421

Qy 1714 GGTGACAAGTGTCTGGACGGCCGTGCTGAACAACAACAGCTACGAGTCAGACATCCTGAAG 1773
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422 GGTGACAAGTGTCTGGACGGCCGTGCTGAACAACAACAGCTACGAGTCAGACATCCTGAAG 481

Qy 1774 AATTACCTGGCAACCAGAGGTTTGACATGGAAAAAACATGTTGACCGAGAGCCTTCGTGGCT 1833
         ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   482 AATTACCTGGCAACCAGAGGTTTGACATGGAAAAAACATGTTGACCGAGAGCCTTCGTGGCT 541

Qy 1834 CTCCAGCGGGGAGTGTTTCTGCGTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGT 1893
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   542 CTCCAGCGGGGAGTGTTTCTGCGTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGT 601

Qy 1894 TACTGCTGTGGCCTGGCGCAGCTTCCGTGAGCTGACCTATCAGTCGGCAGACATTCCT 1953
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   602 TACTGCTGTGGCCTGGCGCAGCTTCCGTGAGCTGACCTATCAGTCGGCAGACATTCCT 661

Qy 1954 GCTTCCGAGTTGCCAGTGGCCAGTGGCGTAACATCCCGTCTACTGGGGCCGTAACTGC 2013
         |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   662 GCTTCCGAGTTGCCAGTGGCCAGTGGCGTAACATCCCGTCTACTGGGGCCGTAACTGC 721

Qy 2014 CGCACTCA-GGTGAAAGCTC-ACCACGCCATGAAATTCAATCATATCTGT-GAACAGACA 2070
         ||||||| |||||||||||| ||||||||||||||||||||||||||||| ||||||||
Db   722 GCACTCAGGGTGAAAGCTCAACCAGCCCATGAAATTCAATCATATCTGTGGGACCAGGAC 781

Qy 2071 AGGTTCAAAAACTAAGCATCCAGAGGCCCTGAG 2103
         ||||||| | ||||||||||||||||||||

source   1..938
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6374110"
                /tissue_type="large cell carcinoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH_MGC_18"
                /note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GGCACGAG(G). Library constructed by Ling Hong in the
                laboratory of Gerald M. Rubin (University of California,
                Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                Superscript II RT (Life Technologies). Note: this is a
                NIH_MGC Library."

ORIGIN

Query Match
Best Local Similarity   26.7%;   Score 714; DB 13;   Length 938;
Matches 874; Conservative 89.7%; Pred. No. 1.3e-108;
          Conservative 0; Mismatches 47; Indels 53; Gaps 8;

Qy   48 CAGCCGGTTCCGGGTTCGGGTTCGGCGGCCGCCGGGGGGCGGGGATGGAATCCCGATGGAGCGGCCCGAGGA 107
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CAGCCGGTTCCGGGTTCGGGTTCGGCGGCCGCCGGGGGGCGGGGATGGAATCCCGATGGAGCGGCCCGAGGA 60

Qy   108 AGGCAAGCAGTCGCCGCCGCCGCCGCCGCCGCCCGCTCTGGGACCGGCTCTCGGCGTCTGGGCGGCGCGAGGA 167
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AGGCAAGCAGTCGCCGCCGCCGCCGCCGCCGCCCGCTCTGGGACCGGCTCTCGGCGTCTGGGCGGCGCGAGGA 120

Qy   168 GGGGAGCGAGCCGCACGTCCTCCTGGAGGAAGCGGAGTCGGACCATCGGCGGCGGGAGACGAGGTTG 227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GGGGAGCGAGCCGCACGTCCTCCTGGAGGAAGCGGAGTCGGACCATCGGCGGCGGGAGACGAGGTTG 180

Qy   228 CGACCTTCCTTCCCCCAGCAATAAACTGGTCTCTGGAAGATCACCAGCAGTGAACAGTGATTAACAA 287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CGACCTTCCTTCCCCCAGCAATAAACTGGTCTCTGGAAGATCACCAGCAGTGAACAGTGATTAACAA 240

Qy   288 TGAAAAATCAGTCAGTGACACTGGAAGATACCAGTGGAACAGTGATTAACAA 347
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TGAAAAATCAGTCAGTGACACTGGAAGATACCAGTGGAACAGTGATTAACAA 300

Qy   348 GCTGAAGGTTGTTAAGAAGCAGACAGACATGCCCTTTACAGACTGGGGATGTCATCATCACTTGGT 407
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GCTGAAGGTTGTTAAGAAGCAGACAGACATGCCCTTTACAGACTGGGGATGTCATCATCACTTGGT 360

Qy   408 GTACAGGAAGAATGACCGAACCAACACAACGTGGCATACCTCTATGAATCTTTAAGTGAAAA 467
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GTACAGGAAGAATGACCGAACCAACACAACGTGGCATACCTCTATGAATCTTTAAGTGAAAA 420

Qy   468 GCAAGGCATGACAACAAGAATCCTTGAAGCTAACAAGGAAAATGTGTTCCATGGGACCAA 527
          |||||||||||||||||||||||||||||||||||||
Db   421 GCAAGGCATGACAACAAGAATCCTTT------------------------------------ 445

Qy   528 AGATACCTCAGGTGCAGGTGCAGGTGCCGAGGGCCGATCCCGGGTCCCTCCGTCGTCGCC 587
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   446 -GATACCTCAGGTGCAGGTGCAGGTGCCGAGGGCCGATCCCGGGTCCCTCCGTCGTCGCC 504

Qy   588 CGCCACTCAGGTGTGCTTTGAGGAACCACCAGCCATCAACATCGACGTCAGACCCTTCCC 647
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   505 CGCCACTCAGGTGTGCTTTGAGGAACCACCAGCCATCAACATCGACGTCAGACCCTTCCC 564

Qy   648 CACAGCCTCGGCCTCTTCCTCCACGGAGCCTTCCTCCGCAGGGGCGAGAGCGTTCCTCCAGTTG 707
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   565 CACAGCCTCGGCCTCTTCCTCCACGGAGCCTTCCTCCGCAGGGGCGAGAGCGTTCCTCCAGTTG 624

Qy   708 TGGGTCTGGGGGTGGTTGGCATCTCCCTAAA-GGAAGTGGTCCCTCGTGTGGCAAGTGATG 766
          |||||||||||||||||||||||||||||||||
Db   625 TGGGTCTGGGGGTGGTTGGCATCTCCCTAAAGGGAAGTGGTCCCTCGTGTGGCAAGTGATG 684

Qy   767 AAGTCTCCAGCTTTGCCTCAGCTCTCCCAGACACAGAAGACTGCGTCCTTTTCGTCGTTGG 826
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   685 AAGTCTCCAGCTTTGCCTCAGCTCTCCCAGACACAGAAGACTGCGTCCTTTTCGTCGTTGG 744

Qy   345 CAAGCTGAAGGTTGTTAAGAAGCAGACAGACATGCCCTTTACAGACTGGGGATGTCATCATCACTT 404
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CAAGCTGAAGGTTGTTAAGAAGCAGACAGACATGCCCTTTACAGACTGGGGATGTCATCATCACTT 360

Qy   405 GGTGTACAGGAAGAATGAACCGAACCAACACAACGTGGCATACCTCTATGAATCTTTAAGTGAA 464
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GGTGTACAGGAAGAATGAACCGAACACAACACAACGTGGCATACCTCTATGAATCTTTAAGTGAA 420

Qy   465 AAAGCAAGGCATGACAACAAGAATCCTTTGAAGCTAACAAGGAAAATGTGTTCCATGGGAC 524
          |||||||||||||||||
Db   421 AAAGCAAGGCATGACAACAAGAATCCTTT------------------------------ 448

Qy   525 CAAAGATACCTCAGGTGCAGGTGCAGGTGCCGAGGGCCGATCCCGGGTCCCTCCGTCGTC 584
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   449 ---GATACCTCAGGTGCAGGTGCAGGTGCCGAGGGCCGAGGGCCGATCCCCGGTCCCTCCGTCGTC 504

Qy   585 GCCCGCCACTCAGGTGTGCTTTGAGGAACCACCAGCCATCAACATCGACGTCAGACCCTTT 644
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   505 GCCCGCCACTCAGGTGTGCTTTGAGGAACCACCAGCCATCAACATCGACGTCAGACCCTTT 564

Qy   645 CCCCACAGCCTCGGCCTCTTCCACGCGAGCCTTCTCCTGCAGGGCGAGAGCGGTTCCTCCAG 704
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   565 CCCCACAGCCTCCGGCCTCTTCCACGCGAGCCTTCTCCTGCAGGGCGAGAGCGGTTCCTCCAG 624

Qy   555 TTGTGGGTCT-GGGGGTGTGGCATCTCCCCTAAAGGAAGTGGTCCCTCGTGTGGCAAGTG 763
          ||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   705 TTGTGGGTCTGGGGGGTGTGGCATCTCCCCTAAAGGAAGTGGTCCCTCGTGTGGCAAGTG 684

Qy   764 ATGAAGTCTCCAGTTTGCCTCAGCTCTCCCAGACACAGAAAGACTGCGTCC-TTTTCGTCG 822
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   685 ATGAAGTCTCCAGTTTGCCTCAGCTCTCCCAGACACAGAAAGACTGCGTCCTTTTTCGTCG 744

Qy   823 TTGGAACCCCAGATC--AGGAGGATTTGGAGCCCGTGAAGAAGAAAATGAGGAGGATG 880
          ||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   745 TTGGAACCCCAGGATCCAGGAGGATTTGGAGCCCGTGAAGAAGAAAATGANAGGAGGATG 804

Qy   881 GGGACCTT--GACCTGAACGGGCAGTTGTTGG---TCGGCACAACCGCGTAGAAAATGCCCA 935
          |||||||| |||||||||||||||||||||| |||||||||||||||||||
Db   805 GGGACCTTTGACCTGGACGGGCAGTTGTGGTGGGGCCGTGGGTCGCACAAACCGCCCTAGAAAATGCCCC 864

Qy   936 AACCG 940
          |||||
Db   865 AACCG 869

RESULT 13
BQ924736
LOCUS   BQ924736   8853777 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6374110
DEFINITION AGENCOURT_8853777 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6374110
            938 bp   mRNA   linear   EST 20-AUG-2002
            5', mRNA sequence.
ACCESSION   BQ924736
VERSION   BQ924736.1   GI:22339767
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 938)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2552 row: i column: 23
            High quality sequence stop: 629.
FEATURES   Location/Qualifiers

Qy   1870 GTCACGGAGGAGACACCGTTCTGTGTTACTGTGGCTGCGCAGCTTCCGTGAGCTGACC 1929
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 GTCACGGAGGAGACACCGTTCTGTGTTACTGTGGCTGCGCAGCTTCCGTGAGCTGAC 121

Qy   1930 TATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCCGTCCT 1989
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 TATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCCGTCCT 181

Qy   1990 GACTGCTACTGGGGCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCCATGAAATTC 2049
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 GACTGCTACTGGGGCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCCATGAAATTC 241

Qy   2050 AATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGCCCTGAGCAGCTT 2109
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 AATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGCCCTGAGCAGCTT 301

Qy   2110 TCAGCACTGGAGGTGAAGAGAGCGGTGTTTTTAAAATACAGAGACAAACACGTCAAGGTGT 2169
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 TCAGCACTGGAGGTGAAGAGAGCGTGTTTTTAAAATACAGAGACAAACACGTCAAGGTGT 361

Qy   2170 TTTCACAGCCCCTGAGGGAAGGACGCAGGGTCTCCGACAGGTGCTCTGGGGTGACTCT 2229
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362 TTTCACAGCCCCTGAGGGAAGGACGCAGGGTCTCCGACAGGTGCTCTGGGGTGACTCT 421

Qy   2230 TCTGTGGAGCTTTTACCCTCGAGTGAGACCCTCCCGAGCCCCGGGGGCCGCAGCCC 2289
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422 TCTGTGGAGC--TTTACCCTCGAGTGAGACCCTCCCGAGCCCCGGGGGCCGCAG-CC 478

Qy   2290 GCCCTCCTGGTGAGCGCTGGGCAGGGCTCGTGGTGGCATCAGCAGCAGCAGAGCAAGCCTT 2349
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   479 GCCCTCCTGGTGAGCGCTGGGCAGGGCTCGTGGTGGCATCAGCAGCAGCAGAGCAAGCCTT 538

Qy   2350 TCTGTAACATGCGGCCGTCCCGCCAGAGAGGGCAGTTTGCTCTTTTGTACATTTTCCGA 2409
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   539 TCTGTAACATGCGGCCGTCCCGCCAGAGAGGGCAGTTTGCTCTTTTGTACATTTTCCGA 598

Qy   2410 AACTACAGTTAAAGCAGAAGTCTGTTTTCAGG-AAAAGTTTCAAGGAGGAAGGGCAAGTT 2468
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   599 AACTACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAAAGTTTCAAGGAGGAAGGGCAAGTT 658

Qy   2469 TATCAAAAACATTGTTTCAGGAGAAGGAGCCATAAGTTTACAGCTAC-AGGACGTACAC 2527
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   659 TATCAAAAACA-TGTTTCCGGAGAAGGAGCCATAAGTTTACAGCGTTACAGACGTACAC 717

Qy   2528 AATATCCTGCTGCTGGGAAAACCACAGCATTTATTCTATTTTTTATTTTTATAGGTTTGG 2587
           ||||||||||||||||||||||||||||||||||||||| ||||| |||||||
Db   718 AATATCCTGCTGCTGGGAAAACCACAGCATTTAATCTATTTTTATCTAATA---GGTTGG 773

Qy   2588 TGCTTATCTTCTAATAAGATTAAAATGTCACAAACTGTAGCACAAATAAT 2637
           ||||||||||||||||||||||||||| ||||||||||||||||
Db   774 GGCTTATCTTCTAATACGATTAAAATGTCACCAAACCTGTAGCCAAATATT 823

RESULT 15
BX409790
LOCUS   BX409790   944 bp   mRNA   linear   EST 13-MAY-2003
DEFINITION BX409790 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF017YI19 5-PRIME, mRNA sequence.
ACCESSION   BX409790
VERSION   BX409790.1 GI:30656944
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 944)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Poleyes,D.
  TITLE   Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Qy   827 AA-CCCCAGGATCAGGAGGGATTTGGAGCCC-----GTGAAGAAGAAAATGAGAGAGATG 880
           || ||||||||||||||||||||||||||| |||||||||||||
Db   745 AACCCCCAGGAATCAGGAGGGATTTGGGAGCCCCNGGAAGAAAAAAAATGANNAGAAGATGGG 804

Qy   881 GGGACCTTGACCTGAA-----CGGGCAGTTGTTGGTCGCACAACCGCGTAGAAATGCCCA 935
           |||||||||||||| |||||||||||||||||||||||||||||||||
Db   805 GGGACCCTTGACCTGAAACACGGGCAGTTGTTTGGTCGCACCAACGCCGTAGGAAATGCCCA 864

Qy   936 AACCGTCCACGAGGA---CGTCAGAGCAGCGGCGGCTGGGAAG-CCAGACAAGAT-GGAGGAG 990
           |||||||||||||| ||||||||||||||||||||| |||||||||| ||||||
Db   865 AACCGTCCACGAGGAACGCCAGAAGCAGCCCAGACCCGCTGGGAAGCCCAGACAAGATGGGGAGGAG 924

Qy   991 ACGCTGACATGCAT 1004
           ||||||||||||||
Db   925 ACGCTGACATGCAT 938

RESULT 14
BG831289
LOCUS   BG831289   824 bp   mRNA   linear   EST 22-MAY-2001
DEFINITION 602766120F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4908316 5',
            mRNA sequence.
ACCESSION   BG831289
VERSION   BG831289.1 GI:14178876
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 824)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1814 row: g column: 05
            High quality sequence stop: 745.
FEATURES   Location/Qualifiers
     source   1..824
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4908316"
                     /tissue_type="epithelioid carcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_42"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by Ling
                     Hong in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

Query Match   26.4%;   Score 708.4; DB 12;   Length 824;
Best Local Similarity 95.7%;   Pred. No. 1.1e-107;
Matches 794; Conservative   0; Mismatches   26; Indels   10; Gaps   6;

Qy   1810 ATGTTGACCGAGAGCCTCGTGGCTTCCACGGCGGGGAGTGTTTCTGCTGCTGCTGATTACAGA 1869
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 ATGTTGACCGAGAGCCTCGTGGCTTCCACGGCGGGGAGTGTTTCTGCTGCTGCTGATTACAGA   61

Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6792.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=S0BAF014ZC07_AF01300_1&cluster=6792.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0BAF014ZC07_AF01300_1.

FEATURES   Location/Qualifiers
     source   1..944
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="CS0DF017YI19"
     /tissue_type="FETAL BRAIN"
     /dev_stage="fetal"
     /clone_lib="Homo sapiens FETAL BRAIN"
     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
     was primed with a NotI-oligo(dT) primer. Five prime end
     enriched, double-strand cDNA was digested with Not I and
     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
     vector. Library was not normalized."

ORIGIN

Query Match   26.4%;   Score 706.2;   DB 13;   Length 944;
Best Local Similarity   89.7%;   Pred. No. 2.5e-107;
Matches   849;   Conservative   0;   Mismatches   84;   Indels   13;   Gaps   8;

```
Qy   1578  AGCGGAGCGCGAGCAGGACCCGCGTGTCGCCCTCAGCAGTGTGCGGTCTGCCTGCAGCC  1637
           ||||||||||||||||||||||||||  |||||||||||| ||||||||||||||||||
Db      4  AGCGGAGCGCGAGCAGGACCCGCGTGTCGCCCCT-AGCAGTGTGCGGTCTGCCTGCAGCC    62

Qy   1638  TTTCTGCCACCCTGTACTGGGGCTGCACCCGACCGGCTGCTACGGCTGCCTGCCCCGTT  1697
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63  TTTCTGCCACCCTGTACTGGGGCTGCACCCGACCGGCTGCTACGGCTGCCTGCCCCGTT   122

Qy   1698  TTGTGAGTCAACCTGGGTGACAAGTGTCTGGACGGCGTGCTGAACAACAACAGCTACGA  1757
           ||||||                    ||  |||||||||||||||||||||||||||||
Db    123  TTGTGAATCGTAATGAAAAGAAATATGTTTATCCTCAGTGTCGGGCAGACAGTGCTTTC   182

Qy   1758  GTCAGACATCCTGA-----AGAATTACCTGGCAACCAGGAGTTTGACATGGAAAAACATG  1812
           ||||| ||| ||||     ||||||||||||||||||||||||||||||||||||||||
Db    183  CTGCGACATCATTATAAACAGAATTACCTGGCAACCAGGAGTTTGACATGGAAAAACATG   242

Qy   1813  TTGACCGAGAGCCCTCGTGGCTCTTCCAGCGGGGAGTGTTTCTGCTGTCTGATTACAGAGTC  1872
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243  TTGACCGAGAGCCCTCGTGGCTCTTCCAGCGGGGAGTGTTTCTGCTGTCTGATTACAGAGTC   302

Qy   1873  ACGGGAGACACCCGTTCTGTTGTACTGCTGTGGGCCTGCGCCAGCTTCCGTGAGTGACCTAT  1932
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303  ACGGGAGACACCCGTTCTGTTGTACTGCTGTGGGCCTGCGCCAGCTTCCGTGAGTGACCTAT   362

Qy   1933  CAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCTGAC  1992
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363  CAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCTGAC   422

Qy   1993  TGCTACTCGGGGCCGTAACTGCCGACTCAGGTGAAAGCTCACCACGCCATGAAATTCAAT  2052
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423  TGCTACTCGGGGCCGTAACTGCCGACTCAGGTGAAAGCTCACCACGCCATGAAATTCAAT   482

Qy   2053  CATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCA  2112
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    483  CATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCA   542

Qy   2113  GCACTGGAGGTGAAGAGAGAGCGTGTTTTAAAATACAGAGACAAGCACGTCAAGTGTTTT  2172
           |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    543  GCACTGGAGGTGAAGAGAGAGCGTGTTTTAAAATACAGAGACAAGCACGTCAAGTGTNTNT   602

Qy   2173  CACAGCCCCCTGAGGGAAGGGACCGCAGGGTCTCCGACACAGGTGCTCTGGGTGACTCTTCT  2232
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603  CACACCCCCTGAGGGAAGGGACCGCAGGGGTCTCCACAGTGCTCTGGGGTGACTCTTCT   662

Qy   2233  GTGGAGCTTTTTACCCTCGAGTGAGACCCTCCCCAGAGCCCTCCCCAGAGCCCCAGCCCC   2292
```

```
Db    663  GTGGAGC--TTTACCCTCTGAGTGAGACCCTCCCCAGAG-CCCGGGGGGCCGCAGACGAGCGAGCCCGCC   719
           ||||||  ||||||||||||||||||||||||||||||  ||| ||||||||||||||||||||||||||
Qy   2293  CTCCTGGTGAGCGCTGGGCAGGGCTCGTGGTGGCATCAGCAGCAGCAGAGACGAGAGCCTTTCT  2352
           |||||||||||||||||||||||||||||||||||||||| ||||||  ||||||
Db    720  CTCCT-GTGAGCGCTGGGCAGGGCTCGTGGTGGCATTAGCAGCAAAAACGAAG-CTTTCT   777
Qy   2353  GTAACATGCGGCCGTCCCGCCGAGAGGGGCAGTTTTGCTCTTTTGTAC-ATTTTCCGAAA  2411
           |||||||||||||||||||||||||||||||||||||||||||||||  ||||||||
Db    778  GTAACATGCGGCCGTTCCGCCGAGAAGGGCAGTTTTGGTTCTTTGGACAATTTTCCGAAC   837
Qy   2412  CTACAGTTAAA-GCAGAAGTCTGTTTCAGGAAAAGTTTCAGGGAGAAGGGCAAGTTTA  2470
           |||| |||||| |||||  ||  ||||||||||  ||||||||||||||||||||||
Db    838  CTACCGTTAAAGGCGAAGGTTGGTTAAGAAAAGTGTTAAGGGAAAAAGGGCCAGTTT   897
Qy   2471  TCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTAC  2516
           ||||||||||||||||| ||||| |||||||||| ||||
Db    898  ATAAAAACATTTGTTAAGGAAAGGGAGCCCTTAGTTTTTCAGTCCAC   943
```

Search completed: May 14, 2004, 20:06:21
Job time : 10045 secs

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Run on: May 14, 2004, 20:10:20 ; Search time 16426 Seconds
                  (without alignments)
                  7069.036 Million cell updates/sec

Title: US-10-048-046-1
Perfect score: 2679
Sequence: 1 aagattcggcacgaggccg.........acaaaaaaaaaaaaaaa 2679

Scoring table: OLIGO_NUC
                  Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 2167151699S residues

Word size :   0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :   GenEmbl:*
          1: gb_ba:*
          2: gb_htg:*
          3: gb_in:*
          4: gb_om:*
          5: gb_ov:*
          6: gb_pat:*
          7: gb_ph:*
          8: gb_pl:*
          9: gb_pr:*
         10: gb_ro:*
         11: gb_sts:*
         12: gb_sy:*
         13: gb_un:*
         14: gb_vi:*
         15: em_ba:*
         16: em_fun:*
         17: em_hum:*
         18: em_in:*
         19: em_mu:*
         20: em_om:*
         21: em_or:*
         22: em_ov:*
         23: em_pat:*
         24: em_ph:*
         25: em_pl:*
         26: em_ro:*
         27: em_sts:*
         28: em_un:*
         29: em_vi:*
         30: em_htg_hum:*
         31: em_htg_inv:*
         32: em_htg_other:*
         33: em_htg_mus:*
         34: em_htg_pln:*
         35: em_htg_rod:*
         36: em_htg_mam:*
         37: em_htg_vrt:*
         38: em_sy:*
         39: em_htgo_hum:*
         40: em_htgo_mus:*
         41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Query Length	DB	ID	Description
	1	2679	100.0	2679	9	AF170724	AF170724 Homo sapi
	2	2133	79.6	2639	6	AX877194	AX877194 Sequence
	3	2133	79.6	2639	6	BD156534	BD156534 Primer fo
	4	2133	79.6	2639	9	AK027687	AK027687 Homo sapi
	5	2031	75.8	3189	9	BC012072	BC012072 Homo sapi
	6	1618	60.4	3181	9	AX405642	AX405642 Sequence
	7	1497	55.9	2297	9	HSM802304	AL137561 Homo sapi
	8	1465	54.7	3138	6	AX877222	AX877222 Sequence
	9	1465	54.7	3138	6	BD156548	BD156548 Primer fo
	10	1465	54.7	3138	6	AK001658	AK001658 Homo sapi
	11	944	35.2	2448	6	AX834949	AX834949 Sequence
	12	944	35.2	2448	6	AK097671	AK097671 Homo sapi
	13	930	34.7	6235	6	BD183291	BD183291 Novel gen
	14	909	33.9	2257	9	AK090948	AK090948 Homo sapi
	15	633	23.6	2186	9	AX713476	AX713476 Sequence
	16	633	23.6	2186	6	AK054917	AK054917 Homo sapi
	17	478	17.8	816	6	AX868758	AX868758 Sequence
	18	478	17.8	816	6	BD148820	BD148820 Primer fo
	19	468	17.5	824	6	AX868776	AX868776 Sequence
	20	468	17.5	824	6	BD148838	BD148838 Primer fo
	21	426	15.9	518	6	AX873789	AX873789 Sequence
c	22	426	15.9	518	6	BD153851	BD153851 Primer fo
c	23	386	14.4	181012	9	AC127070	AC127070 Homo sapi
	24	386	14.4	181438	2	AC023047	AC023047 Homo sapi
c	25	218	8.1	181438	2	AC023047	AC023047 Homo sapi
c	26	78	2.9	617	6	AX385952	AX385952 Sequence
	27	38	1.4	158897	2	AC118260	AC118260 Mus muscu
	28	37	1.4	3194	10	BC049792	BC049792 Mus muscu
	29	37	1.4	321708	2	AC087142	AC087142 Mus muscu
	30	37	1.4	338116	2	AC087159	AC087159 Mus muscu
c	31	34	1.3	202792	2	AC125898	AC125898 Rattus no
	32	32	1.2	601	11	BV002855	BV002855 S20P6639
	33	29	1.1	125396	2	AC124951	AC124951 Medicago
c	34	28	1.0	146516	2	AC079981	AC079981 Homo sapi
	35	28	1.0	218226	2	AC126634	AC126634 Rattus no
c	36	28	1.0	228987	2	AC111838	AC111838 Rattus no
c	37	28	1.0	274796	2	AC097665	AC097665 Rattus no
	38	27	1.0	167166	2	AC117568	AC117568 Mus muscu
	39	27	1.0	167904	2	AC115904	AC115904 Mus muscu
c	40	27	1.0	217842	2	AL929073	AL929073 Mus muscu
c	41	27	1.0	230221	10	AC123851	AC123851 Mus muscu
c	42	27	1.0	234393	10	AL772311	AL772311 Mouse DNA
	43	27	1.0	236825	2	AC140355	AC140355 Mus muscu
	44	27	1.0	239783	2	AC102564	AC102564 Mus muscu
c	45	27	1.0	266632	2	AC129762	AC129762 Rattus no

ALIGNMENTS

RESULT 1
AF170724
LOCUS   AF170724   2679 bp   mRNA   linear   PRI 03-AUG-2000
DEFINITION   Homo sapiens cell cycle checkpoint protein CHFR mRNA, complete cds.
ACCESSION   AF170724
VERSION   AF170724.1 GI:9651169
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2679)
AUTHORS   Scolnick,D.M. and Halazonetis,T.D.
TITLE   Chfr defines a mitotic stress checkpoint that delays entry into
             metaphase

Pred. No. is the number of results predicted by chance to have a

```
Qy  541 GCAGGTGCAGGGCGAGGGGCGATCCCGGTCCCTCCGTCGCCGCACTCAGGTG 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 GCAGGTGCAGGGCGAGGGGCGATCCCGGTCCCTCCGTCGCCGCACTCAGGTG 600

Qy  601 TGCTTTGAGGAACCACAGCCATCAACATGACGTCAGACCTCTTCCCACAGCTCGGCC 660
        |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db  601 TGCTTTGAGGAACCACAGCCATCAACATCACGTCAGACCTCTTCCCACAGCTCGGCC 660

Qy  661 TCTTCCACGGAGCCTTCTCCTGCGAGGGCGAGAGCGGTTCCTCCCAGTGTGGGTCTGGGGGT 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 TCTTCCACGGAGCCTTCTCCTGCGAGGGCGAGAGCGGTTCCTCCCAGTGTGGGTCTGGGGGT 720

Qy  721 GGTGGCATCTCCCCTAAAGGAAGTGGTGTCCCTCGTGTGGCAAGTGATGAAGTCTCCAGCTTT 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 GGTGGCATCTCCCCTAAAGGAAGTGGTGTCCCTCGTGTGGCAAGTGATGAAGTCTCCAGCTTT 780

Qy  781 GCCTCAGCTCTCCAGACAGAAAGACTGCGTCCTTTCGTCGTTGGAACCCCAGGATCAG 840
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 GCCTCAGCTCTCCAGACAGAAAGACTGCGTCCTTTCGTTCGTTGGAACCCCAGGATCAG 840

Qy  841 GAGGATTTGGAGCCCGTGAAGAAGAAATGAGAGAGATGGGACCTTGACCTGAACGGG 900
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 GAGGATTTGGAGCCCGTGAAGAAGAAATGAGAGAGATGGGACCTTGACCTGAACGGG 900

Qy  901 CAGTTGTTGGTCGCACAACCGGTAGAAATGCCAAACCGTCCACGAGACGTCAGAGCA 960
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 CAGTTGTTGGTCGCACAACCGGTAGAAATGCCAAACCGTCCACGAGACGTCAGAGCA 960

Qy  961 GCGCGTGGGAAGCCAGAACAAGATGAGGAGGAACGCGTGCATGCATCATCCTGCCAGGACCTG 1020
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 GCGCGTGGGAAGCCAGAACAAGATGAGGAGGAACGCGTGCATGCATCATCCTGCCAGGACCTG 1020

Qy 1021 CTGCACGACTGCGTGAGTTTGCAGCCCTGCTTCTCGCGGCTTGCTCGCGGTGTACTCG 1080
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 CTGCACGACTGCGTGAGTTTGCAGCCCTGCTTCTCGCGGCTTGCTCGCGGTGTACTCG 1080

Qy 1081 GGCTGGATGGAGCGCTTGTCCGTCCAACACCTGCACCTGCCGTGGAGCGGATCTGT 1140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 GGCTGGATGGAGCGCTTGTCCGTCCAACACCTGCACCTGCCGTGGAGCGGATCTGT 1140

Qy 1141 AAAAACCACATCCTCAACACCTCGTGGAAGCATACCTCATCCAGCATCCAGACAAGAGT 1200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 AAAAACCACATCCTCAACACCTCGTGGAAGCATACCTCATCCAGCATCCAGACAAGAGT 1200

Qy 1201 CGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTG 1260
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 CGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTG 1260

Qy 1261 CAGCCCAAAGTCAGGCGGTCTTTTCGATGAAGAAGGGAGTTCAGAGGACCTGCTGGAG 1320
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 CAGCCCAAAGTCAGGCGGTCTTTTCGATGAAGAAGGGAGTTCAGAGGACCTGCTGGAG 1320

Qy 1321 CTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCAGCCATACGTGTGTGCGGCAG 1380
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 CTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCAGCCATACGTGTGTGCGGCAG 1380

Qy 1381 TGTCCTGAGTACAGAGGCAGGCAGGCGGCGCAGCCTCCCACTGCCCAGCACCGGAGGGCGAG 1440
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 TGTCCTGAGTACAGAGGCAGGCAGGCGGCGCAGCCTCCCACTGCCCAGCACCGGAGGGCGAG 1440

Qy 1441 CCAGGAGCCCCACAGGCCCTGGGGATGCCTGCCCTCTGCGAAGGAGCCACGCACCTCGACGACGAAT 1500
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 CCAGGAGCCCCACAGGCCCTGGGGATGCCTGCCCTCTGCGAAGGAGCCACGCACCTCGACGACGAAT 1500

Qy 1501 GTCCAGGATTACGTGTGCCCTTGCACCTGCTTCCAG 1560
        ||||||||||||||||||||||||||||||||||||
Db 1501 GTCCAGGATTACGTGTGCCCTTGCACCTGCTTCCAG 1560

Qy 1561 CCCATGCCCGACCGGAGGCCGAGAGGCGGAGGCGAGGACCCGGGTGCCCCTCAGCAGTGT 1620
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 CCCATGCCCGACCGGAGGCCGAGAGGCGGAGGCGAGGACCCGGGTGCCCCTCAGCAGTGT 1620

Qy 1621 GCGGTCTGCCTGCCCTTCTGCCACCTGTACTGGGGCTGCACCCGGACCGGCTGCTAC 1680
```

FEATURES   Location/Qualifiers
source   1..2679
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
CDS   91..2085
           /note="contains FHA domain and ring finger"
           /codon_start=1
           /product="cell cycle checkpoint protein CHFR"
           /protein_id="AAF91084.1"
           /db_xref="GI:9651170"
           /translation="MERPEEGKQSPPPQPWGRLLRLGAEBEGEPHVLLRKREWTIGRRR
           GCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDV
           IYLVYRKNEPEHNVAYLYESLSEKQGMTQESPEANKENVFHGTKDTSGAGRGADPR
           VPPSSPATQVCFEREPQPSTSTSDLFPTASASSTEPSPAGRERSSSCGSGGGISPKGS
           GPSVASDEVSFASALPDRKTASFSSLEPDQDEDLEPVKKKMRGDGLDLLNGQLLVAQ
           PRRNAQTVHEDVRAAAGKPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWME
           RSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQP
           KVRRSFSDEEGSSEDLLELSDVDSBSSDISQPYVVCRQCPEYRRQAAQPHCPAPEGE
           PGAPQALGDAPSTSVSLITTAVQDYVCPLQGSHALCTCFQPMPDRRAEREQDPRVAPQ
           QCAVCLQPFCHLYWGCTRTGCYGLCLAPFCELNLGDKCLDGVLNNSYESDILKNYLAT
           RGLTWKNMLTESLVALQRGVFLLSYRVTGDTVLCYCCGLRSFRELTYQRQNIPASE
           LPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQTRFKN"
ORIGIN

Query Match   100.0%; Score 2679; DB 9; Length 2679;
Best Local Similarity 100.0%;
Matches 2679; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 AAGAATTCGGCACGAGGCCGCAATGTCTTGACAGCGGCGCGGCGCGGAGGCCAGCAGCTCCGG 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AAGAATTCGGCACGAGGCCGCAATGTCTCTTGACAGCGGCGCGGCGCGGAGGCCAGCAGCTCCGG 60

Qy   61 GTTCGGCGGCGGGGGCGGGGATGTGAATCCCGATGGAGCGGCGCCCGAGGAAGGCAAGCAGTCG 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GTTCGGCGGCGGGGGCGGGGATGTGAATCCCGATGGAGCGGCGCCCGAGGAAGGCAAGCAGTCG 120

Qy  121 CCGCCGCCGCAGCCCTGGGGACGGCTCCTGCGCTCTGGGCGCGCGGGCGGAGGAGGGCGAGCCCGCAC 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CCGCCGCCGCAGCCCTGGGGACGGCTCCTGCGCTCTGGGCGCGCGGGCGGAGGAGGGCGAGCCCGCAC 180

Qy  181 GTCCTCCTGAGGAAGGCGGGAGTTGGACCATCGGGCGGGAGACAGAGGTTGCGACCTTTCCTTC 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GTCCTCCTGAGGAAGGCGGGAGTTGGACCATCGGGCGGGAGACAGAGGTTGCGACCTTTCCTTC 240

Qy  241 CCCAGCAATAAACTGGTCTCTGGAGATCACTGTGAGAATTGTAGTGAGTGGATGAAAAATCAGGT 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 CCCAGCAATAAACTGGTCTCTGGAGATCACTGTGAGAATTGTAGTGAGTGGATGAAAAATCAGGT 300

Qy  301 CAGTTGACACTGAAGATACCACCAGTGACCACCAAGCACCAGTGAACAGCTGATTAACAAGCTGAAGGTTGTT 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CAGGTGACACTGAAGATACCACCAGTGACCACCAAGCACCAGTGAACAGCTGATTAACAAGCTGAAGGTTGTT 360

Qy  361 AAGAAGCAGACATGCCCTTTACAGACTGGGAATGTCATCATCTTTAAGTGTACAGGAAGAAT 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AAGAAGCAGACATGCCCTTTACAGACTGGGAATGTCATCATCTTTAAGTGTACAGGAAGAAT 420

Qy  421 GAACCGGAACACAAGCGTGCACAACGTGACATACCTCCTCAAATCTTTAAGTGAAAAGCAAGGCATGACA 480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 GAACCGGAACACAAGCGTGCACAACGTGACATACCTCCTCAAATCTTTAAGTGAAAAGCAAGGCATGACA 480

Qy  481 CAAGAATCCTTTGAAGCTAACAGGAGAAAATGTGTTCCATGGGACCAAAGATACCTCAGGT 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 CAAGAATCCTTTGAAGCTAACAGGAGAAAATGTGTTCCATGGGACCAAAGATACCTCAGGT 540
```

```
AX877194
LOCUS       AX877194              2639 bp    DNA      linear   PAT 17-DEC-2003
DEFINITION  Sequence 12099 from Patent EP1074617.
ACCESSION   AX877194
VERSION     AX877194.1  GI:40031930
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primers for synthesising full-length cDNA and their use
  JOURNAL   Patent: EP 1074617-A 12099 07-FEB-2001;
            Research Association for Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1..2639
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             79..2037
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAE89614.1"
                     /db_xref="GI:40031931"
                     /translation="MERPEEGKQSPPPQWGRLLRLGAEBGEPHVLLRKREWTIGRRR
                     GCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSGTVINKLKVVKKQTCPLQTGDV
                     IYLVRKNEPEHNVAYLYESLSEKOGMTQESFDTTSGAGAGRGADPRVPSSPATQVCF
                     EEPQPSTSTSDLFPTASASTEPSPAGRERRSSCGSGSGGGISPKGSGPSVASDEVSSF
                     ASALFDRKTASFSSLEPQDQEDLEPVKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDV
                     RAAAGKPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTRCPV
                     ERICKNHILNNLVEAYLICHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGS
                     SEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPGALGDAPS
                     TSVSLTTAVQDYVCPLQGSHALCTCCFQMPDRRAEREQDPRVAPQQCAVCLQPFCHL
                     YWGCTRTGCYGCLAPFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTES
                     LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCY
                     WGRNCRTQVKAHHAMKFNHICEQTRFKN"
ORIGIN

Query Match         79.6%;  Score 2133;  DB 6;  Length 2639;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 2133;      Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Db   1621 GCGGTCTGCTGCCTCAGCCTTTCTGCTGCCACCTGTACTGGGGCTGCTGTAC 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   1681 GGCTGCCTGGCCCCGTTTTGTGAGCTCAACCTGGTGACAAGTGTCTGGACGGCGTGCTG 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GGCTGCCTGGCCCCGTTTTGTGAGCTCAACCTGGTGACAAGTGTCTGGACGGCGTGCTG 1740

Qy   1741 AACACAACAGCTACAGAGTCAGACATCCTGAAGAATTACCTGCAACCAGAGGTTTGACA 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AACACAACAGCTACAGAGTCAGACATCCTGAAGAATTACCTGCAACCAGAGGTTTGACA 1800

Qy   1801 TGGAAAAACATGTTGACCGAGAGCCTCGTGGCCTTCCAGCGGGGAGTGTTTCTGCTGTCT 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGGAAAAACATGTTGACCGAGAGCCTCGTGGCCTTCCAGCGGGGAGTGTTTCTGCTGTCT 1860

Qy   1861 GATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGTGGCCTGCGCAGCTTCCGT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGTGGCCTGCGCAGCTTCCGT 1920

Qy   1921 GAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACA 1980

Qy   1981 TCCCGTCCTGACTGCTACTGCGGGCCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCC 2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 TCCCGTCCTGACTGCTACTGCGGGCCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCC 2040

Qy   2041 ATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGGCCT 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 ATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGGCCT 2100

Qy   2101 GAGCAGCTTTCAGCACTGAGAGTGAAGAGAGCGCGTGTTTAAAATACAGAGACAAGCACG 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GAGCAGCTTTCAGCACTGAGAGTGAAGAGAGCGCGTGTTTAAAATACAGAGACAAGCACG 2160

Qy   2161 TCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGACGCAGGGTCCCGACAGTGCTCTGG 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGACGCAGGGTCCCGACAGTGCTCTGG 2220

Qy   2221 GGTGACTCTTCTGTGGAGCTTTTACCCTCTGAGTGAGACCCTCCCCCAGAGCCCCGGGGG 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GGTGACTCTTCTGTGGAGCTTTTACCCTCTGAGTGAGACCCTCCCCCAGAGCCCCGGGGG 2280

Qy   2281 CCGCAGCCGCCGCCCTCCTGGTGACGCGCTGGGCAGGGCTCGTGGTGGCATCAGCAGCAGAGA 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CCGCAGCCGCCGCCCTCCTGGTGACGCGCTGGGCAGGGCTCGTGGTGGCATCAGCAGCAGAGA 2340

Qy   2341 CGAAGCCTTTCTGTAACATGCGGCCGTCCCGAGAGGGGCCAGTTTTGCTCTTTTGTAC 2400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CGAAGCCTTTCTGTAACATGCGGCCGTCCCGAGAGGGGCCAGTTTTGCTCTTTTGTAC 2400

Qy   2401 ATTTTCCGAAACTACAGTTAAAGCAGAGAGCAGAAGTCTGTTTCAGGGAAAAGTTTCAAGGGAGAAG 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 ATTTTCCGAAACTACAGTTAAAGCAGAGAGCAGAAGTCTGTTTCAGGGAAAAGTTTCAAGGGAGAAG 2460

Qy   2461 GGCAAGTTTATCAAAAACATTGTTTCAGGAGAGGGAGCATAAGTTACAGCCTACAGGA 2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GGCAAGTTTATCAAAAACATTGTTTCAGGAGAGGGAGCATAAGTTACAGCCTACAGGA 2520

Qy   2521 CGTACACAATAATCCTGCTGCTGGGAAAAACCACCAGCATTTTATCTATTTTTATTTTATTA 2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 CGTACACAATAATCCTGCTGCTGGGAAAAACCACCAGCATTTTATCTATTTTTATTTTATTA 2580

Qy   2581 GGTTTGGTGCTTATCTTCTAATAGATTTAAATGTCACACAACTGTAGCACAAATAATAATA 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 GGTTTGGTGCTTATCTTCTAATAGATTTAAATGTCACACAACTGTAGCACAAATAATAATA 2640

Qy   2641 ATTTATAATTTACAAATTGACAAAAAAAAAAAAAAAA 2679
          |||||||||||||||||||||||||||||||||||||
Db   2641 ATTTATAATTTACAAATTGACAAAAAAAAAAAAAAAA 2679
```

```
Query Match         79.6%;  Score 2133;  DB 6;  Length 2639;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 2133;      Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    529 GATACCTCAGGTGCAGGTGCAGGGGCAGAGGGCGAGGGGCGATCCCCGGTCCCTCCGTCGTCGCCC 588
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GATACCTCAGGTGCAGGTGCAGGGGCAGAGGGCGAGGGGCGATCCCCGGTCCCTCCGTCGTCGCCC 540

Qy    589 GCCACTCAGGTGCGTTGAGGAACCACCAGCCATCAACATGACGTCAGACCTTCCTCCCC 648
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCCACTCAGGTGCGTTGAGGAACCACCAGCCATCAACATGACGTCAGACCTTCCTCCCC 600

Qy    649 ACAGCCTCGGCCTCTTCCTGCAGGGCCTTCTCCTGCAGGGCGGCAGAGCGTTCCTCCCAGTTGT 708
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ACAGCCTCGGCCTCTTCCTGCAGGGCCTTCTCCTGCAGGGCGGCAGAGCGTTCCTCCCAGTTGT 660

Qy    709 GGGTCTGGGGTGGTGGCATCCCGTGGCAATCTCCCTAAAGGAAGTGGTCCCTCGTTGTCGTTGGAA 768
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GGGTCTGGGGTGGTGGCATCCCGTGGCAATCTCCCTAAAGGAAGTGGTCCCTCGTTGTCGTTGGAA 720

Qy    769 GTCTCCAGCTTTGCCTCAGCTCCCAGCTCTCCCAGCACAGCAGAAGACTGCGCGTCCTTTCGTCGTTGGAA 828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GTCTCCAGCTTTGCCTCAGCTCCCAGCTCTCCCAGCACAGCAGAAGACTGCGCGTCCTTTCGTCGTTGGAA 780

Qy    829 CCCCAGGATCAGGAGGATTTGAGCCCGTGAGAGAAGAAATGAGAGGAGATGGGGACCTT 888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CCCCAGGATCAGGAGGATTTGAGCCCGTGAGAGAAGAAATGAGAGGAGATGGGGACCTT 840

Qy    889 GACCTGAACGGCCAGTTGTTGGTCGCACAACCGCTGAAATGCCAAACCGTCCACGAG 948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GACCTGAACGGCCAGTTGTTGGTCGCACAACCGCTGAAATGCCAAACCGTCCACGAG 900

Qy    949 GACGTCAGACGCAGCGGCGGCTGGAAGCCAGCCAGACAAGATGAGGACGCTGACAATGACATCATC 1008
```

RESULT 2

```
Db  1981 GCTCACCACGGCCATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCA 2040
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2089 TCCAGAGGGCCCTGAGCAGCTTTCAGCACTGGAGCTGAGGTGAGAGAGAGCGTGTTTTAAAATACA 2148
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 TCCAGAGGGCCCTGAGCAGCTTTCAGCACTGGAGCTGAGGTGAGAGAGAGCGTGTTTTAAAATACA 2100

Qy  2149 GAGACAAGCAGTCAAGGTGTTTCACAGCCCTGAGGGAAGGGACGCAGGGTCTCCGA 2208
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 GAGACAAGCACAGTCAAGGTGTTTCACAGCCCTGAGGGAAGGGACGCAGGGTCTCCGA 2160

Qy  2209 CAGGTGCTCTGGGGTGACTCTTCTGTGGAGCTTTTTACCCTCTGAGTGAGACCCTCCCA 2268
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 CAGGTGCTCTGGGGTGACTCTTCTGTGGAGCTTTTTACCCTCTGAGTGAGACCCTCCCA 2220

Qy  2269 GAGCCCGGGGCCGCAGCCCGCCCCTCCTGGTGAGCGGCTCGGCGGGCTCGTGGTGGTGCAT 2328
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 GAGCCCCGGGGCCGCAGCCCGCCCCTCCTGGTGAGCGGCTCGGCGGGCTCGTGGTGGTGCAT 2280

Qy  2329 CAGCAGCAGAGACGAGAGCCTTTCGTAACATGCGGCCGTCCCGCCGAGAGGGGCAGTTTT 2388
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 CAGCAGCAGAGACGAGAGCCTTTCGTAACATGCGGCCGTCCCGCCGAGAGGGGCAGTTTT 2340

Qy  2389 GCTCTTTTGTACATTTTCGAAACTACAGTAAAGCAGAAGTCTGTTTTCAGGAAAAGTT 2448
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 GCTCTTTTGTACATTTTCGAAACTACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAGTT 2400

Qy  2449 TCAAGGAGAAGGCAAGTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTA 2508
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 TCAAGGAGAAGGCAAGTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTA 2460

Qy  2509 CAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTT 2568
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 CAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTT 2520

Qy  2569 TTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAAATGTCACAAACTGTAGC 2628
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 TTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAAATGTCACAAACTGTAGC 2580

Qy  2629 ACAAATATATAATTATAATTTACAAATTGAC 2661
         |||||||||||||||||||||||||||||||
Db  2581 ACAAATATATAATTATAATTTACAAATTGAC 2613
```

RESULT 3

```
BD156534
LOCUS       BD156534                2639 bp    DNA     linear   PAT 17-JAN-2003
DEFINITION  Primer for synthesizing full-length cDNA and use thereof.
ACCESSION   BD156534
VERSION     BD156534.1  GI:27862292
KEYWORDS    JP 2002191363-A/11377.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2639)
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primer for synthesizing full-length cDNA and use thereof
  JOURNAL   Patent: JP 2002191363-A 11377 09-JUL-2002;
            HELIX RESEARCH INSTITUTE
COMMENT     OS   Homo sapiens (human)
            PN   JP 2002191363-A/11377
            PD   09-JUL-2002
            PF   28-JUL-2000 JP 2000280990
            PI   TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
            PI   SAITO,
            PI   JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
            PI   KEIICHI NAGAI,TETSUJI OTSUKI
            PC   C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
            10,
            PC   C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
            Primer for synthesizing full-length cDNA and use thereof FH Key
```

```
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GACGTCAGAGCAGCAGCGGCTGGGGAAGCCAGACAGATGGAGGACGACGCTGACATGCATATC 960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1009 TGGCCAGGACCTGCTGCACGACTGCGTGAGTTTGCAGCTCGTCCTGCGCGTGTCCCGTG 1068
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TGGCCAGGACCTGCTGCACGACTGCGTGAGTTTGCAGCTCGTCCTGCGCGTGTCCCGTG 1020
Qy  1069 GCTTGCTACTCGGGCTGGATGGAGCGGCTCGGTCGTCGTCCGTGTCCCACTGCTGCTG 1128
Db  1021 GCTTGCTACTCGGGCTGGATGGAGCGGCTCGGTCGTCGTCCGTGTCCCACTGCTGCTG 1080
Qy  1129 GAGCGGATCTGTAAAAACCACATCCTCAACAACCTCTGGGAAGCATACCTCATCCAGCAT 1188
Db  1081 GAGCGGGATCTGTAAAAACCACATCCTCAACAACCTCGTGGAAGCATACCTCATCCAGCAT 1140
Qy  1189 CCAGACAAGAGTGCAGTGAAGAGATGTGCAAAGTATGGATGCCAGAATAAAATCACT 1248
Db  1141 CCAGACAAGAGTGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGAATAAAATCACT 1200
Qy  1249 CAAGACATGTCGCAGCCCAAAGTCAGGCGGTCTTTTCTGATGAAGAAGGAGTTCAGAG 1308
Db  1201 CAAGACATGTCGCAGCCCAAAGTCAGGCGGTCTTTTCTGATGAAGAAGGAGTTCAGAG 1260
Qy  1309 GACCTGCTGGAGCTGTCAGACGTTGCTGACACGTGAGTCCTCAGACATTAGCCAGCCATACGTC 1368
Db  1261 GACCTGCTGGAGCTGTCAGACGTTGCTGACACGTGAGTCCTCAGACATTAGCCAGCCATACGTC 1320
Qy  1369 GTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGCGGCGGCAGCGCGGCCTCCCACTGCCACGA 1428
Db  1321 GTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGCGGCGGCAGCGCGGCCTCCCACTGCCACGA 1380
Qy  1429 CCCGAGGGCGAGCGAGGAGCCCCAGAGCCCTGGGGATGCACCCTCCACGTCGTCAGC 1488
Db  1381 CCCGAGGGCGAGCGAGGAGCCCCAGAGCCCTGGGGATGCACCCTCCACGTCGTCAGC 1440
Qy  1489 CTGACGACGACAGCAGTCCAGGATTACGTGTGCCCTGCCAAGGAGGCCAGCCCGTGTCACC 1548
Db  1441 CTGACGACGACAGCAGTCCAGGATTACGTGTGCCCTGCCAAGGAGGCCAGCCCGTGTCACC 1500
Qy  1549 TGCTGCTTCCAGCCCATGCCCGGTCTGCTCGCAGCCTTTGCGCCAGCCTTTGCGCCAGGAGGCTGCACCACGGG 1608
Db  1501 TGCTGCTTCCAGCCCATGCCCGGTCTGCTCGCAGCCTTTGCGCCAGCCTTTGCGCCAGGAGGCTGCACCACGCC 1560
Qy  1609 CCTCAGCAGTGTGCCGGTCTGCCACGGGTCTGCGAGCCTTTCTGCGACCTGTACTGACGCCACCCGG 1668
Db  1561 CCTCAGCAGTGTGCCGGTCTGCCACGGGTCTGCGAGCCTTTCTGCGACCTGTACTGGGGCTGCACCCGG 1620
Qy  1669 ACCGGCTGCGTCACGGCGGCTGCCTGCCCCGGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTG 1728
Db  1621 ACCGGCTGCGTCACGGCGGCTGCCTGCCCCGGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTG 1680
Qy  1729 GACGGGCGTGCTGAACAACAACAGCTACGACATCCTGAGAGAATTACCTGGCAACC 1788
Db  1681 GACGGGCGTGCTGAACAACAACAGCTACGACATCCTGAGAGAATTACCTGGCAACC 1740
Qy  1789 AGAGGTTTGACATGGAAAAACATGTTGACCGAGCCCTGTGGCTTCTCCAGCGGGGAGTG 1848
Db  1741 AGAGGTTTGACATGGAAAAACATGTTGACCGAGCCCTGTGGCTTCTCCAGCGGGGAGTG 1800
Qy  1849 TTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGTGGCCTG 1908
Db  1801 TTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGTGGCCTG 1860
Qy  1909 CGCAGTTCCGTGAGCTGACCTATCGGCAGAACATTCCTGCTTCCGACACTGCCGCACTCAGTGAA 1968
Db  1861 CGCAGTTCCGTGAGCTGACCTATCGGCAGAACATTCCTGCTTCCGACACTGCCGCACTCAGTGAA 1920
Qy  1969 GTGGCCGTAACATCCCGTCCTACTGGGGCGTAACTGCCGCACTCAGTGAAA 2028
Db  1921 GTGGCCGTAACATCCCGTCCTACTGGGGCGTAACTGCCGCACTCAGTGAAA 1980
Qy  2029 GCTCACCACGGCCATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCA 2088
```

```
FEATURES       FT    CDS    Location/Qualifiers
                     source           (79)..(2034).
                            Location/Qualifiers
                            1..2639
                            /organism="Homo sapiens"
                            /mol_type="genomic DNA"
                            /db_xref="taxon:9606"
ORIGIN

Query Match         79.6%;    Score 2133;   DB 6;   Length 2639;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 2133;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

Qy 529 GATACCTCAGGTGCAGGTGCAGTGCAGGGCGAGGGCCGATCCCCGGGTCCCTCCGTCGTCGCCC 588
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 GATACCTCAGGTGCAGGTGCAGGGGCGAGGGGCCGATCCCCGGGTCCCTCCGTCGTCGCCC 540

Qy 589 GCCACTCAGGTGTGCTTTGAGGAACCACAGCCATCAACATCGACGTCAGACCTCTTCCCC 648
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 GCCACTCAGGTGTGCTTTGAGGAACCACAGCCATCAACATCGACGTCAGACCTCTTCCCC 600

Qy 649 ACAGCCTGGCCTTCCACGGAGCCTTCTCCGCAGGGCGAGAGCCGTTCCTCCAGTTGT 708
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 ACAGCCTCGGCCTTCCACGGAGCCTTCTCCGCAGGGCGAGAGCCGTTCCTCCAGTTGT 660

Qy 709 GGGTCTGGGGGTGGTGCCATCTCCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAA 768
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 661 GGGTCTGGGGGTGGTGCCATCTCCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAA 720

Qy 769 GTCTCCAGCTTTGCCTGCAGCTCTCCAGACAGAAAGACTGCGTCCTTTTCGTCGTTGGAA 828
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 721 GTCTCCAGCTTTGCCTGCAGCTCTCCAGACAGAAAGACTGCGTCCTTTTCGTCGTTGGAA 780

Qy 829 CCCCAGGGATCAGGAGGATTTGGAGCCCGTGAAGAAAGAAAATGAGAGGAGATGGGACCTT 888
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 781 CCCCAGGGATCAGGAGGATTTGGAGCCCGTGAAGAAAGAAAATGAGAGGAGATGGGACCTT 840

Qy 889 GACCTGAACGGGCAGTTGTTGGTCGCACAACCGCGTAGAAATGCCCAAACCGTTCCACGAG 948
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 841 GACCTGAACGGGCAGTTGTTGGTCGCACAACCGCGTAGAAATGCCCAAACCGTTCCACGAG 900

Qy 949 GACCTCAGAGCAGCAGGGCGGCTGGGAAGCCAGACAGAGCCGAGAGCGCTGACATGCATCATC 1008
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 901 GACCTCAGAGCAGCAGGGCGGCTGGGAAGCCAGACAGAGCCGAGAGCGCTGACATGCATCATC 960

Qy 1009 TGGCCAGGAGACTGCTGCACGACTGCCTGCGTGAGTTTGCAGCCCTGCCATGCACAGTTCTGCGCG 1068
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 961 TGGCCAGGAGACTGCTGCACGACTGCCTGCGTGAGTTTGCAGCCCTGCCATGCACAGTTCTGCGCG 1020

Qy 1069 GCTTGCTACTCGGCCTGATGGAGCGCTCGTCCTGCGTCCTTACCTGCCGCGTGTCCCGTG 1128
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 GCTTGCTACTCGGCCTGATGGAGCGCTCGTCCTGCGTCCTTACCTGCCGCGTGTCCCGTG 1080

Qy 1129 GAGCGGATCTGTAAAAACCACATCCTCAACAACTCGTTGGAAGCATACCTCATCCAGCAT 1188
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 GAGCGGATCTGTAAAAACCACATCCTCAACAACTCGTTGGAAGCATACCTCATCCAGCAT 1140

Qy 1189 CCAGACAAGAGTCGCAGTGAACGTGAAGAAGATGTGCAAAGTATGGATGCAAATAAAATCACT 1248
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 CCAGACAAGAGTCGCAGTGAACGTGAAGAAGATGTGCAAAGTATGGATGCAAATAAAATCACT 1200

Qy 1249 CAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTCTGATGAAGAAGGGAGTTCAGAG 1308
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 CAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTCTGATGAAGAAGGGAGTTCAGAG 1260

Qy 1309 GACCTGCTGGAGCTGTCAGACGTTGACAGTGAGTCAGACATTAGCCAGCCATACGTC 1368
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 GACCTGCTGGAGCTGTCAGACGTTGACAGTGAGTCAGACATTAGCCAGCCATACGTC 1320

Qy 1369 GTGTCGGCAGTGTCCTGAGTACAGAAGGCAGGCAGGCGGGGCCCTCCCCACTGCCCAGCA 1428
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 GTGTCGGCAGTGTCCTGAGTACAGAAGGCAGGCAGGCGGGGCCCTCCCCACTGCCCAGCA 1380

Qy 1429 CCCGAGGGGCGGCAGCAGGAGCCCCAGAGCCCTCGGGGGATGCACCTTCCACGTTCCGTCAGC 1488
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 CCCGAGGGGCGGCAGCAGGACCCCAGAGCCCTCGGGGGATGCACCTTCCACGTTCCGTCAGC 1440

Qy 1489 CTGACACAGCAGTCAGATTACGTCGTGCCCTCTGCCAAGGAAGCCACGCCCGTGTGCACC 1548
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 CTGACACAGCAGTCAGATTACGTCGTGCCCTCTGCCAAGGAAGCCACGCCCGTGTGCACC 1500

Qy 1549 TGCTGCTTCAGCCATGCCCGACCGGAGAGCGGAGCCAGGAGCGCGAGCCAGCAGGACCCGCGTGTCGCC 1608
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 TGCTGCTTCAGCCATGCCCGACCGGAGAGCGGAGCCAGGAGCGCGAGCCAGCAGGACCCGCGTGTCGCC 1560

Qy 1609 CCTCAGCAGTGTGCGGTCTGCCTGCAGCCTTTCTGCGCCTTTCTGCGCCTGTACTGGGCGCACCGG 1668
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 CCTCAGCAGTGTGCGGTCTGCCTGCAGCCTTTCTGCGCCTTTCTGCGCCTGTACTGGGCGCACCCGG 1620

Qy 1669 ACCGGCTGCTACGGCTGCCTGGCCCCGTTTTGTGAGCTCAACCTGGTGACAAGTGTCTG 1728
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 ACCGGCTGCTACGGCTGCCTGGCCCCGTTTTGTGAGCTCAACCTGGTGACAAGTGTCTG 1680

Qy 1729 GACGGGGTGCTGAACAACAACAGCTACGAGTCAGAGTCAGACATCCGAAGAATTACCTGGCAACC 1788
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1681 GACGGGGTGCTGAACAACAACAGCTACGAGTCAGAGTCAGACATCCGAAGAATTACCTGGCAACC 1740

Qy 1789 AGAGGTTTGACATGGAAAAACATGTTGACCGGAGAGCCTCGTGGCGTCTCCCAGCGGGGAGTG 1848
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1741 AGAGGTTTGACATGGAAAAACATGTTGACCGGAGAGCCTCGTGGCGTCTCCCAGCGGGGAGTG 1800

Qy 1849 TTTCTGCTGTCTGATTACAGAGTCACGGAGACACCGTTCTGTGTTACTGCTGTGTGGCCTG 1908
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1801 TTTCTGCTGTCTGATTACAGAGTCACGGAGACACCGTTCTGTGTTACTGCTGTGTGGCCTG 1860

Qy 1909 CGCAGCTTCCGTGAGCTGACCTATCAGTAGTATCGGCAGAACATTCCTGCTTCCCAGTTGCCA 1968
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1861 CGCAGCTTCCGTGAGCTGACCTATCAGTAGTATCGGCAGAACATTCCTGCTTCCCAGTTGCCA 1920

Qy 1969 GTGGCCGTAACATCCGTCCTACTGGGGCCGTAACTGCCGCACTCAGTGAAA 2028
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1921 GTGGCCGTAACATCCGTCCTACTGGGGCCGTAACTGCCGCACTCAGTGAAA 1980

Qy 2029 GCTCACCACGCCATGAAATTCAATCATATCTGTGAACAGACAGAAGGTTCAAAAACTAAGCA 2088
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1981 GCTCACCACGCCATGAAATTCAATCATATCTGTGAACAGACAGAAGGTTCAAAAACTAAGCA 2040

Qy 2089 TCCAAGAGGCCCTGACGCAGCTTTCAGCAGCTTTCAGCAGTGTGTTTCAGAGAGTGAAGAGAGCCGGTGTTTTTAAAATACA 2148
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2041 TCCAAGAGGCCCTGACGCAGCTTTCAGCAGCTTTCAGCAGTGTGTTTCAGAGAGTGAAGAGAGCCGGTGTTTTTAAAATACA 2100

Qy 2149 GAGACAAGCACGTCAAGGTGTTTTCACAGCCCCTGAGGGAAGGGACGCAGGGTCTCCGA 2208
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2101 GAGACAAGCACGTCAAGGTGTTTTCACAGCCCCTGAGGGAAGGGACGCAGGGTCTCCGA 2160

Qy 2209 CAGGTGCTGTCGGGGTGACTCTTCTGTGGAGCTTTCTCGTGGCTGGGCTCTGGGTGGTGGCAT 2268
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2161 CAGGTGCTGTCGGGGTGACTCTTCTGTGGAGCTTTTCTCGTGGCTGGGCTCTGGGTGGTGGCAT 2220

Qy 2269 GAGCCCCGGGGCCGCCGCTTTCGTGTAACATGCGGCCCGGTCCCCGAGAGGGCAGTTT 2328
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2221 GAGCCCCGGGGCCGCCGCTTTCGTGTAACATGCGGCCCGGTCCCCGAGAGGGCAGTTT 2280

Qy 2329 CAGCAGCAGAGACGAGCGAAGCGCCCTTTCTGAATACACAGTAAAGCAGAGAAGTTCAGAG 2388
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2281 CAGCAGCAGAGACGAGCGAAGCGCCTTTCTGAATACACAGTAAAGCAGAGAAGGGAGTTCAGAG 2340

Qy 2389 GCTCTTTTGTACATTTTCCGAAACTACAGTTAGCCAGCCATACGTC 2448
        ||||||||||||||||||||||||||||||||||||||||||||||
Db 2341 GCTCTTTTGTACATTTTCCGAAACTACAGTTAGCCAGCCATACGTC 2400

Qy 2449 TCAAGGGAGAAGGGCAAGTTATCAAAAACATTGTTTCAGGAAGAAGGGAGGCATAAGTTTA 2508
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2401 TCAAGGGAGAAGGGCAAGTTATCAAAAACATTGTTTCAGGAAGAAGGGAGGCATAAGTTTA 2460

Qy 2509 CAGCCTACAGGACGTACACAATATCCTGCTGTGGGAAAACCACAGCATTTTATCTATTT 2568

ORIGIN
```
Query Match        79.6%;  Score 2133;  DB 9;  Length 2639;
Best Local Similarity  100.0%;  Pred. No. 0;
Matches 2133;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy   529 GATACCTCAGGTGCAGGTGCAGGGCGAGGGCGCCATCCCCGGGTCCTCCGTCGTCGCCC  588
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GATACCTCAGGTGCAGGTGCAGGGCGAGGGCGCCATCCCCGGGTCCTCCGTCGTCGCCC  540

Qy   589 GCCACTCAGGTGTGCTTTGAGGAACCACAGCCATCAACATCGACTCAGACCTCTTCCCC  648
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GCCACTCAGGTGTGCTTTGAGGAACCACAGCCATCAACATCGACTCAGACCTCTTCCCC  600

Qy   649 ACAGCCTCGGCCTCTTCCACGGAGCCTTCTCCTGCAGGGCGAGAGCGTTCCTCCAGTTGT  708
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ACAGCCTCGGCCTCTTCCACGGAGCCTTCTCCTGCAGGGCGAGAGCGTTCCTCCAGTTGT  660

Qy   709 GGGTCTCTGGGGTGGTGCATCTCCCCTAAAGGAAGTGGTCCCTGTGTGGCAAGTTGATGAA  768
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GGGTCTCTGGGGTGGTGCATCTCCCCTAAAGGAAGTGGTCCCTGTGTGGCAAGTTGATGAA  720

Qy   769 GTCTCCAGCTTGCCTCAGCTCTCCAGACAGCAGAAGACTGCGTCCTTTCGTCGTTGGAA  828
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GTCTCCAGCTTGCCTCAGCTCTCCAGACAGCAGAAGACTGCGTCCTTTCGTCGTTGGAA  780

Qy   829 CCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGACAGAAAATGAGGAGGATGGGGACCTT  888
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 CCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGACAGAAAATGAGGAGGATGGGGACCTT  840

Qy   889 GACCTGAACGGGGCAGTTGTGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAG  948
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 GACCTGAACGGGGCAGTTGTGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAG  900

Qy   949 GACGTCAGGAGCAGCGGCCAGACAGGAAGCCAGACAAGATGGAGGAGACGCTGACATGCATCATC 1008
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GACGTCAGGAGCAGCGGCCAGACAGGAAGCCAGACAAGATGGAGGAGACGCTGACATGCATCATC  960

Qy  1009 TGCCAGGACCTGCTGCACGACTGCGTGAGTTGCAGCCCTGCACACGTTCGCGCG 1068
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TGCCAGGACCTGCTGCACGACTGCGTGAGTTGCAGCCCTGCACACGTTCGCGCG 1020

Qy  1069 GCTTGCTACTCGGGCTGGATGGAGCGGCTCGTCGTCCCGTGTCCTACCTGCCGCTGTCCCGTG 1128
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 GCTTGCTACTCGGGCTGGATGGAGCGGCTCGTCGTCCCGTGTCCTACCTGCCGCTGTCCCGTG 1080

Qy  1129 GAGCGGATCGTGTAAAACCATCCTCAACACCTGTGGAAGCATACCTCATCCAGCAT 1188
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 GAGCGGATCGTGTAAAAACCACATCCTCAACACCTGTGGAAGCATACCTCATCCAGCAT 1140

Qy  1189 CCAGACAAGAGTGCCAGTGAAGAGATGTCAAAGTATGGATGCCAGAATAAATCACT 1248
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CCAGACAAGAGTGCCAGTGAAGAGATGTCAAAGTATGGATGCCAGGAATAAATCACT 1200

Qy  1249 CAAGACAATGCTGCAGCCCAAAGTCAGGCGGGTCTTTTCTGATGAAGAGGAGGTCAGAG 1308
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CAAGACAATGCTGCAGCCCAAAGTCAGGCGGGTCTTTTCTGATGAAGAGGAGGGAGTTCAGAG 1260

Qy  1309 GACCTGCTGGAGCTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTC 1368
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 GACCTGCTGGAGCTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTC 1320

Qy  1369 GTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCGGCGGCCTCCCCACTGCCCAGCA 1428
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 GTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCGGCGGCCTCCCCACTGCCCAGCA 1380

Qy  1429 CCCGAGGGCGGAGCCAGGAGCCCAAAGTGCACCTCCACCCGTCTCAGC 1488
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 CCCGAGGGGCGGAGCCAGGAGCCCACAGGCCCTGGGGGATGCACCCTCCACCCGTCCGTCAGC 1440

Qy  1489 CTGACGACAGCAGTCCAGGATTACGTGTCCCCTGCACCCTGCACC 1548
         |||||||||||||||||||||||||||||||||||||||||||||
Db  1441 CTGACGACAGCAGTCCAGGATTACGTGTCCCCTGCACCCTGCACC 1500

Db  2461 CAGCCTACAGGACGTACACAATATCCTGCTGTTATCTTCTAATAAGATTTAATGTCACAAAACTGTAGC 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2569 TTTATTTTAAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAAACTGTAGC 2628
         ||||||||||||||||||||||||||||||||||||||
Db  2521 TTTATTTTAAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAAACTGTAGC 2580

Qy  2629 ACAAATATATAATTATAATTATAATTACAAATTGAC 2661
         |||||||||||||||||||||||||||||
Db  2581 ACAAATATATAATTATAATTATAATTACAAATTGAC 2613
```

RESULT 4
AK027687
LOCUS   AK027687   2639 bp   mRNA   linear   PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ14781 fis, clone NT2RP4000455, weakly similar
            to TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP0.
ACCESSION   AK027687
VERSION   AK027687.1 GI:14042552
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2639)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,
            Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
            Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2639)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE   Direct Submission
  JOURNAL   Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology; cDNA library construction,
            5'- & 3'-end one pass sequencing and clone selection: Helix
            Research Institute (supported by Japan Key Technology Center etc.)
            and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES   Location/Qualifiers
     source   1..2639
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RP4000455"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2RP4"
                     /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
     CDS   79..2037
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="BAB55297.1"
                     /db_xref="GI:14042553"
                     /translation="MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRR
                     GCDLSFPSNKLVSGDHCRIVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDV
                     IYLVYRKNPEHNVAYLYESLSKQGMTQESPDTSGAGRGADPRVPPSSPATVVCF
                     EEPQPSTSTSDLFPTAGASSTEPSPAGRERSSSCGSGGGGISPKGSGPSVASDEVSSF
                     ASALPDRKTASFSSLEPDDQEDLEPVKKXMRGDLDLLNGQLLVAQPRRNAQTVHEDV
                     RAAAGKPKDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV
                     ERICKNHILNNLVEAYLIQHPDKSRSEBDVQSMDARNKITQDMLQPKVRRSFSDEEGS
                     SEDLLELSDVSSSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGPAGPQALGDAPS
                     TSVSLTTAVQDYVCPLQGSHALCTCCPQPMPDRRABREQDPRVAPQQCAVCLQPPCHL
                     YWGCTRTGCYGCLAPFCBLNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTES
                     LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQRQNIPASELPVAVTSRPDCY
                     WGRNCRTQVKAHHAMKFNHICEQTRFKN"

```
Qy   2629 ACAAATATATATAATTATAATTTACAAATTGAC 2661
          |||||||||||||||||| ||||||||||||||
Db   2581 ACAAATATATATAATTTATAATTTACAAATTGAC 2613
```

```
RESULT 5
BC012072
LOCUS       BC012072                3189 bp    mRNA    linear   PRI 04-OCT-2003
DEFINITION  Homo sapiens checkpoint with forkhead and ring finger domains, mRNA
            (cDNA clone MGC:19963 IMAGE:4650348), complete cds.
ACCESSION   BC012072
VERSION     BC012072.1  GI:15082329
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3189)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmitz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 3189)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Pjel, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 29 Row: n Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 8922674.
FEATURES             Location/Qualifiers
     source          1..3189
```

```
Qy   1549 TGCTGCTTCCAGCCATGCCCGACCGGAGAGAGGCGGGACGCAGGACCCGCGTGTCGCC 1608
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TGCTGCTTCCAGCCATGCCCGACCGGAGAGAGGCGGGACGCAGGACCCGCGTGTCGCC 1560

Qy   1609 CCTCAGCAGTGTGCGGTCGCCTGCCTTCTGCCACCTGTACTGGGCTGCACCCCG 1668
          ||||||||||||||||||| ||||||||||||||||||||||||||||| |||||
Db   1561 CCTCAGCAGTGTGCGGTCTGCCTGCCTTCTGCCACCTGTACTGGGCTGCGACCCCG 1620

Qy   1669 ACCGGCTGCTACGGCTGCCTGCCCGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTG 1728
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ACCGGCTGCTACGGCTGCCTGCCCGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTG 1680

Qy   1729 GACGGGCGTGCTGAACAACAACAGCTACGAGTCAGACATCCTGAAGAATTACCTGCAACC 1788
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GACGGGCGTGCTGAACAACAACAGCTACGAGTCAGACATCCTGAAGAATTACCTGCAACC 1740

Qy   1789 AGAGGTTTGACATGGAAAAACATGTGACCGGAGAGCCTCGTGGCTCTCCAGCGCGGGAGTG 1848
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AGAGGTTTGACATGGAAAAACATGTGACCGGAGAGCCTCGTGGCTCTCCAGCGCGGGAGTG 1800

Qy   1849 TTTCTGCTGTGATTACAGAGTCACGGGAGACACCGTTCGTGTTACTGCTGTGGCCTG 1908
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TTTCTGCTGTGATTACAGAGTCACGGGAGACACCGTTCGTGTTACTGCTGTGGCCTG 1860

Qy   1909 CGCAGCTTCCGTGAGCTGACCTTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCA 1968
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CGCAGCTTCCGTGAGCTGACCTTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCA 1920

Qy   1969 GTGGCCGTAACATCCCGTCCGTACTGGGGCCGTAACTGCCGCACTCAGGTGAAA 2028
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GTGGCCGTAACATCCCGTCCGTACTGGGGCCGTAACTGCCGCACTCAGGTGAAA 1980

Qy   2029 GCTCACCACGCCCATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCA 2088
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GCTCACCACGCCCATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCA 2040

Qy   2089 TCCAGAGGCCCTGAGCAGCTTTCAGCACTGAGGTGAAGAGAGCGTGTTTTTAAAATACA 2148
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   2041 TCCAGAGGCCCTGAGCAGCTTTCAGCACTGAGGTGAAGAGAGCGTGTTTTTAAAATACA 2100

Qy   2149 GAGACAAGCACGTCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGGACGCAGGGTCTCCGA 2208
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GAGACAAGCACGTCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGGACGCAGGGTCTCCGA 2160

Qy   2209 CAGGTGCTCTGGGGTGACTCTTCTGTGGAGCTTTTACCCTCTGAGTGAGACCCTCCCCA 2268
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CAGGTGCTCTGGGGTGACTCTTCTGTGGAGCTTTTACCCTCTGAGTGAGACCCTCCCCA 2220

Qy   2269 GAGCCCCGGGGGGCCGCAGCCCGCCCCTCCTGGGCTGAGCGCTCGGGCAGGGCTCGTGGTGGCAT 2328
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GAGCCCCGGGGGGCCGCAGCCCGCCCCTCCTGGGCTGAGCGCTCGGGCAGGGCTCGTGGTGGCAT 2280

Qy   2329 CAGCAGCAGAGACGAAGCCTTTCTGTAACATGCGGCGTCCCGCCGAGAGGGGCAGTTTT 2388
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CAGCAGCAGAGACGAAGCCTTTCTGTAACATGCGGCGTCCCGCCGAGAGGGGCAGTTTT 2340

Qy   2389 GCTCTTTTGTACATTTTCCGAAACTACAGTAAAGCAGAAGTCTGTTTCAGGAAAAGTT 2448
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GCTCTTTTTGTACATTTTCCGAAACTACAGTAAAGCAGAAGTCTGTTTCAGGAAAAGTT 2400

Qy   2449 TCAAGGGAGAAGGGCAAGTTTATCAAAACATTGTTTCAGGAGGAGGGAGCATAAGTTTA 2508
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 TCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGGAGGGAGCATAAGTTTA 2460

Qy   2509 CAGCCTACAGAGACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTT 2568
          |||||||||||||||||| |||||||||||||||||||||||||||||||||| |||||||
Db   2461 CAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTT 2520

Qy   2569 TTTATTTTAATAGGTTTGGTGCTTATCTTCTCAATAAGATTTAAATGTCACAACTGTAGC 2628
          |||||||||| || |||||||||||||||||||||||||| ||||||||||||||||||
Db   2521 TTTATTTTAATAGGTTTGGTGCTTATCTTCTCAATAAGATTTAAATGTCACAAACTGTAGC 2580
```

```
Db   771 GACCTGAACGGCGAGTTGTTGGTCGCCACAACCGCGTAGAAATGCCCAAACCGTCCACGCAG 830

Qy   949 GACGTCAGAGCAGCGGCTGGGAAGCCAGACAAGATGGAGAGACGCTGACATGCATCATC 1008
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   831 GACGTCAGAGCAGCAGCGGCTGGGAAGCCAGACAAGATGGAGAGACGCTGACATGCATCATC 890

Qy  1009 TGCCAGGACCTGCTGCACGACTGCGTGAGTTTGCAGCCCGTTGCAGCACACGTTCTGCGCG 1068
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   891 TGCCAGGACCTGCTGCACGACTGCGTGAGTTTGCAGCCCGTTGCAGCACACGTTCTGCGCG 950

Qy  1069 GCTTGCTACTCGGGCGGCTGGATGGAGCGCTCGTCCTGTGTCCTACTCCTGCCGTCCCGTG 1128
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   951 GCTTGCTACTCGGGCGGCTGGATGGAGCGCTCGTCCTGTGTCCTACTCCTGCCGTCCCGTG 1010

Qy  1129 GAGCGGATCTGTAAAAACCACCATCCTCAACAACCTCGTGGAAGCATACCTCATCCAGCAT 1188
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1011 GAGCGGATCTGTAAAAACCACCATCCTCAACAACCTCGTGGAAGCATACCTCATCCAGCAT 1070

Qy  1189 CCAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAGTATGGATGCGAAGTATGGATAAAATCACT 1248
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1071 CCAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAGTATGGATGCGAAGTATGGATAAAATCACT 1130

Qy  1249 CAAGACATGCTGCAGCCAAAGTCAGCGGTCTTTTCTGATGAAGAAGGGAGTTCAGAG 1308
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1131 CAAGACATGCTGCAGCCAAAGTCAGCGGTCTTTTCTGATGAAGAAGGGAGTTCAGAG 1190

Qy  1309 GACCCTGCTGAGCGTCTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTC 1368
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1191 GACCTGCTGGGAGCTGTCAGACGTTGCAGTGAGTCCTCAGACATTAGCCAGCCATACGTC 1250

Qy  1369 GTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCGGGCGGAGCCTCCCACTGCCCAGCA 1428
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1251 GTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCGGGCGGAGCCTCCCACTGCCCAGCA 1310

Qy  1429 CCCGAGGGGCGAGCCAGAGCCCACGCACCTCCAGTCCGTCAGC 1488
         |||||||||||||||||||||||||||||||||||||||||||
Db  1311 CCCGAGGGGCGAGCCAGAGCCCACGCACCTCCAGTCCGTCAGC 1370

Qy  1489 CTGACGACAGCAGTCCAGGATTACGTGTGCCCTGCCAAGGAAGCCACGCCCGTGTCACC 1548
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1371 CTGACGACAGCAGTCCAGGATTACGTGTGCCCTGCCAAGGAAGCCACGCCCGTGTCACC 1430

Qy  1549 TGCTGCTTCCAGCCCATGCGGCGAGCGGAGCGCGGAGCGGAGCCGGAGACCGCGTGTCGCC 1608
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1431 TGCTGCTTCCAGCCCATGCGGCGAGCGGAGCGGAGCGGAGCCGGAGACCGCGTGTCGCC 1490

Qy  1609 CCTCAGCAGTGTGCGGTCTGCGCTGCTGCCTGGCTGCGCACCGTGTACTGGGGGCTCACCCGG 1668
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1491 CCTCAGCAGTGTGCGGTCTGCGCTGCTGCCTGGCTGCGCACCGTGTACTGGGGGCTGCACCCGG 1550

Qy  1669 ACCGGCTGCTACGGCCTGCTGCCTGGCTGCCATCAACATGCAGACCTCAGACCTCTTCCCC 1728
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1551 ACCGGCTGCTACGGCCTGCTGCCTGGCTGCCATCAACATGCAGACCTCAGACCTCTTCCCC 1610

Qy  1729 GACGGGCGTGCTGAACACAACAGCTCAGTCAGTACGACATCGCAGAATTACCTGCAACC 1788
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1611 GACGGGCGTGCTGAACACAACAGCTCAGTCAGTACGACATCGCAGAATTACCTGCAACC 1670

Qy  1789 AGAGGTTTGACATGGAAAAACATGTGACCGAGAGCCTCGTGGCCTCCAGCAAGTGATGAA 1848
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1671 AGAGGTTTGACATGGAAAAACATGTGACCGAGAGCCTCGTGGCCTCCAGCAAGTGATGAA 1730

Qy  1849 TTTCTGCTGTCTGATTACAGAGTCAGGGAGACACCGTTCTGCTTTTCGTCGTTGGAA 1908
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1731 TTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGCTTTTCGTCGTTGGAA 1790

Qy  1909 CGCAGCTTCCGTGAGCTGACCTATCAGTAGTCGGCAGAACATTCCTGCTTCCAGTTGCCA 1968
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1791 CGGAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCAGTTGCCA 1850

Qy  1969 GTGGCCGTAACATCCCGTCCACTGCTACTGCGGGCGCACTCAGGTGAAA 2028
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  1851 GTGGCCGTAACATCCCGTCCACTGCTACTGCGGGCGCACTCAGGTGAAA 1910
```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:19963 IMAGE:4650348"
/tissue_type="Placenta, choriocarcinoma"
/clone_lib="NIH MGC_21"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1..3189

gene
/gene="CHFR"
/note="synonyms: FLJ10796, RNF116"
/db_xref="LocusID:55743"
/db_xref="MIM:605209"

CDS
9..1967
/codon_start=1
/product="CHFR protein"
/protein_id="AAH12072.1"
/db_xref="GI:15082330"
/db_xref="LocusID:55743"
/translation="MERPEBGKQSPPPQPWGRLLRLGAEEGEPHVLLRKQREWTIGRRR
GCDLSFPSNKLVSGDHCRIVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDV
IYLVYRKNEPEHNVAYLYESLSEKQGMTQESFDTSGAGRGADPRVPPSSPATQVCF
EEPQPSTSTSDLFPTASASSTEPSPAGRERRSSCSGGGISPKGSGPSVASDEVSSF
ASALPDRKTASFSSLEPQDQEDLEPVKKKVRGDGDLDLNGQLLVAQPRRNAQTVHEDV
ERICKNHILNNLVEAYLIQHPDKSRSREEDVQSMDARNKITQDMLQPKVRRSFSDEEGS
SEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPBGEPGAPQALGDAPS
TSVSLTTAVQDYVCPLQGSHALCTCCFQPMEDRRVEREBQDPRVAPQQCAVCLQPFCHL
YWGCTRTGCYGCLAPFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTES
LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCY
WGRNCRTQVKAHHAMKFNHICEBQTRFKN"

misc_feature
120..317
/note="FHA; Region: FHA domain. The FHA
(Forkhead-associated) domain is a phosphopeptide binding
motif"
/db_xref="CDD:pfam00498"

misc_feature
879..1010
/note="HRD1; Region: COG5243, HRD1, HRD ubiquitin ligase
complex, ER membrane component [Posttranslational
modification, protein turnover, chaperones]"
/db_xref="CDD:COG5243"

ORIGIN

Query Match   75.8%;   Score 2031;   DB 9;   Length 3189;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches 2131;   Conservative 0;   Mismatches 0;   Indels 2;   Gaps 0;

```
Qy   529 GATACCTCAGGTGCAGGTGCAGGGCGAGGGCCGATCCCCGGTCCCTCCGTCGTCGCCC 588
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   411 GATACCTCAGGTGCAGGTGCAGGGCGAGGGCCGATCCCCGGTCCCTCCGTCGTCGCCC 470

Qy   589 GCCACTCAGGTGTGCTTTGAGGAACCAGACCAGCCATCAACATCGACGTCAGACCTCTTCCCC 648
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   471 GCCACTCAGGTGTGCTTTGAGGAACCAGACCAGCCATCAACATCGACGTCAGACCTCTTCCCC 530

Qy   649 ACAGCCTGGCCTTCCACGGAGGCCTTCTCCTGCAGGGCGGAGAGCGTTCCTCCAGTTGT 708
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   531 ACAGCCTCGGCCTTCCACGGAGGCCTTCTCCTGCAGGGCGGAGAGCGTTCCTCCAGTTGT 590

Qy   709 GGGTCTCGGGGGTGGTGGCATCTCCCCTAAAGGAAGTGGTCCCTGTGTGGCAAGTGATGAA 768
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   591 GGGTCTGCGGGGGTGGTGGCATCTCCCCTAAAGGAAGTGGTCCCTGTGTGGCAAGTGATGAA 650

Qy   769 GTCTCCAGCTTGCCTCAGCTCTCCGAGGAACTGCCTCCTTTTCGTCGTTGGAA 828
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   651 GTCTCCAGCTTGCCTCAGCTCTCCGAGGAACTGCCTCCTTTTCGTCGTTGGAA 710

Qy   829 CCCCAGGATCAGAGAGGATTTGGAGCCCGTGAACAGAGAAATGAGAGGACTT 888
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   711 CCCCAGGATCAGAGAGGATTTGGAGCCCGTGAACAGAGAAATGAGAGGACTT 770

Qy   889 GACCTGAACGGCGAGTTGTTGGTCGCCACAACCGCGTAGAAATGCCCAAACCGTCCACGCAG 948
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
/db_xref="REMTREMBL:CAD34749"
/translation="MERPEGKQSPPPQPWGRLLRLRGABEGEPHVLLRKREWTIGRRR
GCDLSFPSNKLVSGDHCRIVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDV
IYLVYRKNEPEHNVAYLYESLSEKQGMTQESFEMVPCCVAQAGLKLLGSDPPTLASQ
SIVITGSGGGISPKGSGPSVASDEVSSPASALPDRKTASFSSLEPQDEDLEPVKKK
MRGDGDLDNGQLLVAQPRRNAQTVHEDVRAAAGKPDKWEETLCIICQDLLHDCVSL
QPCWHTFCAACYSGWMERSSLCP?CRCPVERICKNHILNNLVEAYLIQHPDKSRSEED
VQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPE
YRRQAAQPPHCPAPEGEPGAPQAIGDAPPTSVSLTTAVQDYVCPLQGSHALCTCCFQP
MPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYCLAPPCELNLGDKCLDGV
LNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLR
SFRELTYQYQQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQTRFKN"

ORIGIN
Query Match     60.4%;  Score 1618;  DB 6;  Length 3181;
Best Local Similarity 99.7%;  Pred. No. 0;
Matches 1948;  Conservative  0;  Mismatches  0;  Indels  3;  Gaps  1;
```

Qy   2629 ACAAATAATATAATTTATAATTTACAAATTGAC 2661
           ||||||||||||||||||||||||||||||||
Db   2509 ACAAATAATATAATTTATAATTTACAAATTGAC 2541

RESULT 7
HSM802304
LOCUS   HSM802304   2297 bp   mRNA   linear   PRI 18-FEB-2000
DEFINITION   Homo sapiens mRNA; cDNA DKFZp434N2420 (from clone DKFZp434N2420);
             partial cds.
ACCESSION   AL137561
VERSION   AL137561.1 GI:6808265
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2297)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and
             Wiemann,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152
             Martinsried, GERMANY
COMMENT   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
             Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
             sequenced by MediGenomix (Martinsried/Germany) within the cDNA
             sequencing consortium of the German Genome Project. This clone
             (DKFZp434N2420) is available at the RZPD in Berlin. Please contact
             the RZPD: Resourcenzentrum, Heubnerweg 6, 14059
             Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
             information about the clone and the sequencing project is available
             at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES   Location/Qualifiers
     source   1..2297
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434N2420"
                     /tissue_type="testis"
                     /clone_lib="434 (synonym: htes3). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
     gene   1..2297
                     /gene="DKFZp434N2420"
     CDS   <1..921
                     /gene="DKFZp434N2420"
                     /note="unknown"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB70812.1"
                     /db_xref="GI:6808266"
                     /db_xref="SPTREMBL:Q9NT32"
                     /translation="VEAVLIQHPDKSRBEDVOSMDARNKITODMLQPKVRRSFSDEE
                     GSSEDLLELSDVDSBSSDISQPYVVCRQCPEYRRQAAOPPHCPAPEGEPGAPQALGDA
                     PSTSVSLTTAVODYVCFPLQGSHALCTCCFQPMPDRRAERRQDPRVAPQQCAVCLQPFC
                     HLYWGCTRTGCYGCLAPFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLT
                     ESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPD
                     CYWGRNCRTVKAHHAMKFNHICEQTRFKN"
     polyA_signal   2227..2232
                     /gene="DKFZp434N2420"
     polyA_site   2251
                     /gene="DKFZp434N2420"
ORIGIN

Query Match   55.9%; Score 1497; DB 9; Length 2297;
Best Local Similarity   100.0%; Pred. No. 0;
Matches 1497; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1165 GTGGAAGCATACCTCATCCAGCATCCAGACAAGAGTCGCAGTGAAGAAGATGTGCAAAGT 1224
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTGGAAGCATACCTCATCCAGCATCCAGACAACAAGAGTCGCAGTGAAGAAGATGTGCAAAGT 60

Qy   1225 ATGGATGCCAGAATAAAATCACTCAAGACACATGCTGCCCCAAAGTCAGGCGGGTCTTTT 1284

Qy   1549 TGCTGCTTCCAGCCATGCCGACCCGAGAGGCGGAGCAGCAGACCCGCTGTCGCC 1608
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1431 TGCTGCTTCCAGCCCATGCCCGACCCGAGAGGCGGAGCAGCAGACCCGCTGTCGCC 1490

Qy   1609 CCTCAGCAGAGTGCGGTGCTGCTCGCCGCCTTTCTGCCCACCTGTACTGGGCTGCACCCGG 1668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1491 CCTCAGCAGAGTGCGGTGCTGCTCGCCGCCTTTCTGCCCACCTGTACTGGGCTGCACCCGG 1550

Qy   1669 ACCGGCTGCTACGGCTGCCTGCCTGGCCCCGTTTGTGAGCTGCAACCTGGGTGACAAGTGTCTG 1728
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1551 ACCGGCTGCTACGGCTGCCTGCCTGGCCCCGTTTGTGAGCTGCAACCTGGGTGACAAGTGTCTG 1610

Qy   1729 GACGGCGTGCTGAACAACAACAGCTACGAGTCAGATCAGAGAATTACCTGGCAACC 1788
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1611 GACGGCGTGCTGAACAACAACAGCTACGAGTCAGATCAGAGAATTACCTGGCAACC 1670

Qy   1789 AGAGGTTTGACATGGAAAAACATGTTGACCGAGGCCTCGTGGCTCTCCAGCGGGGAGTG 1848
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1671 AGAGGTTTGACATGGAAAAACATGTTGACCGAGGCCTCGTGGCTCTCCAGCGGGGAGTG 1730

Qy   1849 TTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGTGGCCTG 1908
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1731 TTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGTGGCCTG 1790

Qy   1909 CGCAGCTTCCGTGAGCTGACCTATCAGTAGCGGCAGAACATTCCTGCTTCCGAGTTGCCA 1968
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1791 CGCAGCTTCCGTGAGCTGACCTATCAGTATCAGCAGAACATTCCTGCTTCCGAGTTGCCA 1850

Qy   1969 GTGGCCGTAACATCCCGTCCTGACTGCTACTGGGGCCGTAACTGCCCGCCACTCAGGTGAAA 2028
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1851 GTGGCCGTAACATCCCGTCCTGACTGCTACTGGGGCCGTAACTGCCGCCACTCAGGTGAAA 1910

Qy   2029 GCTCACCACGCCATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAAACTAAGCA 2088
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1911 GCTCACCACGCCATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAAAACTAAGCA 1970

Qy   2089 TCCAGAGGCCCTGAGCAGCTTTCCGCAGCCCGCCGCCCTTCCTGGTGAGCGCGTCGGTGGTGGCAT 2148
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1971 TCCAGAGGCCCTGAGCAGCTTTCCGCAGCCCGCCGCCCTTCCTGGTGAGCGCGTCGGTGGTGGCAT 2030

Qy   2149 GAGACAAGCACGTCAAGGTGTTTCACAGCCCCTGAGGGAGGACGCAGGGTCTCCGA 2208
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2031 GAGACAAGCACGTCAAGGTGTTTCACAGCCCCTGAGGGAGGACGCAGGGTCTCCGA 2090

Qy   2209 CAGTGTCTCTGGGGTGACTCTTCTGTGGAGCTTTTACCCTGAGTGAGACCCTCCCCA 2268
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2091 CAGTGTCTCTGGGGTGACTCTTCTGTGGAGC--TTTACCCTGAGTGAGACCCTCCCCA 2148

Qy   2269 GAGCCCCGGCGGGGCCGCAGCCCGCCCTTCGGTGAGCGCGTCGGTGGTGGCAT 2328
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2149 GAGCCCCGGCGGGGCCGCAGCCCGCCCTTCGGTGAGCGCGTCGGTGGTGGCAT 2208

Qy   2329 CAGCAGCAGAGAGACGAAGCCTTTCGTAACATGCGGCGCGTCCCGCCGAGAGGGGCAGTTT 2388
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2209 CAGCAGCAGAGAGACGAAGCCTTTCGTAACATGCGGCGCGTCCCCGCCGAGAGGGGCAGTTT 2268

Qy   2389 GCTCTTTTGTACATTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAGTT 2448
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2269 GCTCTTTTGTACATTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAGTT 2328

Qy   2449 TCAAGGGAGAAGGGCCAAGTTTATCAAAAACATTGTTCAGGAGAAGGAGCATAAGTTTA 2508
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2329 TCAAGGGAGAAGGGCCAAGTTTATCAAAAACATTGTTCAGGAGAAGGAGCATAAGTTTA 2388

Qy   2509 CAGCCTACAGGACGTACACAATATCCTGCTGCTGGAAAACCACACAGCATTTATCTATTT 2568
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2389 CAGCCTACAGGACGTACACAATATCCTGCTGCTGGAAAACCACACAGCATTTTATCTATTT 2448

Qy   2569 TTTATTTTAATAGGTTTGGTGCTTATCTTCAATAAGATTTAAAGTGCACAAACTGTAGC 2628
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2449 TTTATTTTAATAGGTTTGGTGCTTATCTTCAATAAGATTTAAAGTGCACAAACTGTAGC 2508

RESULT 8

```
AX877222
LOCUS       AX877222                3138 bp    DNA     linear   PAT 17-DEC-2003
DEFINITION  Sequence 12127 from Patent EP1074617.
ACCESSION   AX877222
VERSION     AX877222.1  GI:40031958
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
            Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
  TITLE     Primers for synthesising full-length cDNA and their use
  JOURNAL   Patent: EP 1074617-A 12127 07-FEB-2001;
            Research Association for Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1..3138
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             65..1936
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAB89626.1"
                     /db_xref="GI:40001959"
                     /translation="MERPEEGKQSPPPQPWGRLLRLGAEEGBPHVLLRKREWTIGRRR
                     GCDLSFPSNKLVSGDHCRIVVDEKSGGVTLEDTSTGTVINKLKVVKKQTCPLQTGDV
                     IYLVYRKNEPEHNVAYLYESLSEKQGMTQESFEMVPCCVAQGLKLLGSSDPPTLASQ
                     SIVITSGGGGISPKGSGPSVASDEVSFASALPDRKTASFSSLEPQDQEDLEPEKKK
                     MRGDGDLDLNGQLLVAQPRRNAQTVHBDVRAAAGKPDKMEETLTCIICQDLLHDCVSL
                     QPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEED
                     VQSMDARNKIYQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSSDISQPYVVCRQCPE
                     YRRQAAQPPHCPAPEGBPGAPQALGDAPPTSVSLTTAVQDYVCPLQGSHALCTCCFQP
                     MPDRRARREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELNLGDKCLDGV
                     LNNNSYESDILKNYLATRGLTWKNMLTBSLVALQRGVFLLSDYRVTGDTVLCYCCGLR
                     SFRELTYQYQQNIPASELPVAVTSRPDCYWGRNCRTOVKAHHAMKFNHICEQTRFKN"
ORIGIN
```

Query Match   54.7%; Score 1465; DB 6; Length 3138;
Best Local Similarity   99.6%;
Matches 1945; Conservative   0; Mismatches   0; Indels   2; Gaps   1;

```
PD    09-JUL-2002
PF    28-JUL-2000 JP 200028090
PI    TOSHIO OTA,TAKAO ISOGAI,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
PI    SAITO,
PI    JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI    KEIICHI NAGAI,TETSUJI OTSUKI
PC    C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC
      10,
PC    C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
      Primer for synthesizing full-length cDNA and use thereof FH Key
FT    CDS          Location/Qualifiers
                   1..3138
FEATURES           Location/Qualifiers
      source       1..(1933).
                   /organism="Homo sapiens"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:9606"

ORIGIN

Query Match          54.7%;   Score 1465;   DB 6;   Length 3138;
Best Local Similarity 99.6%;  Pred. No. 0;
Matches 1945;  Conservative 0;  Mismatches 0;  Indels 2;  Gaps 1;
```

```
Qy   709  GGGTCTCGGGGTGGTGGCATCTCCCTAAAGGAAGTGGTGCCCTCTGTTGGCAAGTGATGAA  768
          ||||||||||||||||||||||||||||||||||||||  ||| |||||||||||||||||
Db   560  GGGTCTCGGGGTGGTGGCATCTCCCTAAAGGAAGTGG||CCCTTGTGTGCAAGTGATGAA   619

Db   620  GTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAAGACTGCGGTCCTTTTTCGTTCGTTGAA  679

Qy   829  CCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAAATGAGGAGGATGGGGACCTT        888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   680  CCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAAATGAGGAGGATGGGGACCTT        739

Qy   889  GACCTGAACGGGCAGTTGTTGGTCGCACAACCCGTAGAAATGCCCAAACCGTCCACGAG     948
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   740  GACCTGAACGGGCAGTTGTTGGTCGCACAACCCGTAGAAATGCCCAAACCGTCCACGAG     799

Qy   949  GACGTCAGAGAGCAGCGCGCTGGGAAGCCAGAGCCAGACAAGATGGAGGAGACGCTGACATCATC  1008
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   800  GACGTCAGAGAGCAGCGCGCTGGGAAGCCAGAGCCAGACAAGATGGAGGAGACGCTGACATCATC   859

Qy  1009  TGCCAGGAGACCTGCTGCACGACATCCGTGAGTTTGCAGCCCTGAGCTTTGCACGTTCTGCGGCG  1068
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   860  TGCCAGGAGACCTGCTGCACGACTGCGTGAGTTTGCACGCCCTGAGCTTTGCACGTTCTGCGGCG   919

Qy  1069  GCTTGCTACTCGGGCGATGGAGGAGCGCTCGTCGTTCCTGGTCCTACCTGCCGCGTCGTCCGTG  1128
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   920  GCTTGCTACTCGGGCGATGGAGGAGCGCTCGTCGTTCCTGGTCCTACCTGCCGCGTCGTCCCGGTG   979

Qy  1129  GAGGCGGATCGTAAAAACCACATCCTCAACAACCTCGTGGAAGCATACCTCATCCAGCAT      1188
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   980  GAGGCGGATCGTAAAAACCACATCCTCAACAACCTCGTGGAAGCATACCTCATCCAGCAT      1039

Qy  1189  CCAGACAAGAGTCGCAGTGAAGAAGAGATGTGCAAAGTGCAGGAATATAAAATCACT       1248
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1040  CCAGACAAGAGTCGCAGTGAAGAAGAGATGTGCAAAGTGCAGGAATATAAAATCACT       1099

Qy  1249  CAAGACATGCTGCAGCCCAAAGTCAGGCGGCCCAAAGTTCTTTTTCTGATGAAGAAGAGGAGTTCAGAG  1308
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1100  CAAGACATGCTGCAGCCCAAAGTCAGGCGGCCCAAAGTTCTTTTTCTGATGAAGAAGAGGAGTTCAGAG   1159

Qy  1309  GACCTGCTGGAGCTGTCAGACGTTGACAGTTAGCCAGCCATACGTC      1368
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  1160  GACCTGCTGGAGCTGTCAGACGTTGACAGTTAGCCAGCCATACGTC      1219

Qy  1369  GTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCCAGCCTCCCCACTGCCCAGCA       1428
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1220  GTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCCAGCCTCCCCACTGCCCAGCA       1279

Qy  1429  CCCGAGGGGCGAGCCCAGGAGCCCCACAGGGGGTGGGATGCACCCCTCCACGTCCGTCAGC  1488
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1280  CCCGAGGGGCGAGCCCAGGAGCCCCACAGCCCGGGGATGCCAAGGAGAGCACGCCGCCTGCCACC  1339

Qy  1489  CTGACGACGAGCAGTCCAGGATTACGTGTGCCCTCTGCAAGGAAGCCACGCACCCGTGTGCCACC  1548
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1340  CTGACGACGAGCAGTCCAGGATTACGTGTGCCCTCTGCAAGGAAGCCACGCACCCGTGTGCCACC  1399

Qy  1549  TGCTGCTTCCAGCCATGCCCGACCCGGAGAGCGAGCGGCAGCGGAGCAGAGACCCGGCGTGTCGCC  1608
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1400  TGCTGCTTCCAGCCATGCCCGACCCGGAGAGCGAGCGGCAGCGGAGCAGAGACCCGGCGTGTCGCC  1459

Qy  1609  CCTCAGCAGTGTGCGGTCTGCCTGCAGCCTTTCTGCCACTGTACTGGGGCTGCACCCGG    1668
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1460  CCTCAGCAGTGTGCGGTCTGCCTGCAGCCTTTCTGCCACCGTACTGGGGCTGCACCCGG    1519

Qy  1669  ACCGGCTGCTACGGCTGCCTGGCCCCGTTTTGTGAGCTCACGACTCAACCTGCACTGCAAGTGTCTG  1728
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1520  ACCGGCTGCTACGGCTGCCTGCCCCCGTTTTGTGAGCTCACGACTCAACCTGGACAAGTGTCTG   1579

Qy  1729  GACGGCGTGCTGAACAACAACAGCTACGAGTCAGAGTCAGAATTACTGCAACC      1788
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1580  GACGGCGTGCTGAACAACAACAGCTACGAGTCAGAGTCAGAATTACCTGGCAACC      1639

Qy  1789  AGAGGTTTGACATGGAAAAACATGTGACCGGAGAGCCTCGTGCGCTTCCCAGCGGGGGAGTG   1848
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1640  AGAGGTTTGACATGGAAAAACATGTGACCGGAGAGCCTCGTGCGCTTCCCAGCGGGGGAGTG   1699

Qy  1849  TTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTTCCCTG    1908
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1700  TTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTTCGGCCTG    1759

Qy  1909  CGCAGCTTCCGTGACTGAGCTGACCTATCAGTATCGCAGAACATTCCTGCTTCCGAGTTGCCA  1968
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1760  CGCAGCTTCCGTGACTGAGCTGACCTATCAGTATCAGCAGAACATTCCTGCTTCCGAGTTGCCA  1819

Qy  1969  GTGGCCTAACATCCGTCCTGACTGCTACTGGCCGTAACTGCCCAACCGTCAGGTGAAA      2028
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1820  GTGGCCGTAACATCCGTCCTGACTGCTACTGGGCCGTAACTGCCCAACCGTCAGGTGAAA     1879

Qy  2029  GCTCACCACGCCATGAAATTCAATCATCATATCTGACACGAGCAAGGTTCAAAAACTAAGCA  2088
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1880  GCTCACCACGCCATGAAATTCAATCATCATATCTGACACAGAGCAAGGTTCAAAAACTAAGCA   1939

Qy  2089  TCCAGAGGCCCTGACAGCTTCAGCACTGAGCGTGAAGAGCGTGTTTAAAATACA     2148
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1940  TCCAGAGGCCCTGACAGCTTCAGCACTGAGCGTGAAGAGCGTGTTTTAAAATACA     1999

Qy  2149  GAGACAAGCACGTCAAGGTGTTTCACAGCCCCTGAGGCAAGGGAGACCCAGGGTCTCCGA    2208
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2000  GAGGCAAGCACGTCAAGGTGTTTCACAGCCCCTGAGGCAAGGGAGACCAGGGTCTCCGA    2059

Qy  2209  CAGGTGCTCTGGGGTGACTCTTCTGTGAGCTTTGTCTTCTGTGAGTCAGGCTCGTCGTGGTGCCCCA  2268
          |||||||||||||||||||||||||||||||||||||||  |||||||||||
Db  2060  CAGGTGCTCTGGGGTGACTCTTCTGTGAGC--TTTACCCTGAGTCAGGCTCGTCGTGGTGCCCCA   2117

Qy  2269  GAGCCCCGGGGGGCCGCAGCCCGCAGCCTGGTGTGTAACATGCGGCCGTCCCGAGAGGGGCAGGGCAT  2328
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2118  GAGCCCCGGGGGGCCGCAGCCCGCAGCCTGGTGTAACATGCGGCCGTCCCGAGAGGGGCAGGGCAT   2177

Qy  2329  CAGCAGCAGAGACGAAGCCTTTCGAAACTACAGTAAAGCGAAGTCGTTTTCAGGAAAAGTT   2388
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2178  CAGCAGCAGAGACGAAGCCTTTCGAAACTACAGTAAAGCGAAGTCGTTTTCAGGAAAAGTT   2237

Qy  2389  GCTCTTTTGTACATTTCCGAAACACAGTTTTCAAAAACATTGTTCAGGAGAGGAGCATAAGTTTA   2448
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2238  GCTCTTTTGTACATTTCCGAAACACAGTTTTCAAAAACATTGTTCAGGAGAGGAGCATAAGTTTA   2297

Qy  2449  TCAAGGGAGAAGGCAAGTTTATCAAAAACATTTGTTTATCAAAAACATTGTTCAGGAGAGGAGCATAAGTTTA   2508
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2298  TCAAGGGAGAAGGCAAGTTTATCAAAAACATTTGTTTATCAAAAACATTGTTCAGGAGAAACCAGGAGCATAAGTTTA  2357

Qy  2509  CCCGAGGGGCGAGCCCAGGAGCCCCACAGCCCGGGGATGCACCCCTCCACGTCCGTCAGC   2568
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2509  CAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTT
```

```
Query Match         54.7%;  Score 1465;  DB 9;  Length 3138;
Best Local Similarity  99.6%;  Pred. No. 0;
Matches 1945;  Conservative  0;  Mismatches  0;  Indels  6;  Gaps  1;

Qy    709 GGGTCTGGGGTGGTGGCATCTCCCCTAAAGGAAGTGGTCCCTCTGGCAAGTGATGAA 768
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    560 GGGTCTGGGGTGGTGGCATCTCCCCTAAAGGAAGTGGTCCCTCTGGCAAGTGATGAA 619

Qy    769 GTCTCCAGCTTGCCTCAGCTCTCCAGACAGCAGAAAGACTGCGTCCTTTCGTCGTTGGAA 828
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    620 GTCTCCAGCTTGCCTCAGCTCTCCAGACAGCAGAAAGACTGCGTCCTTTCGTCGTTGGAA 679

Qy    829 CCCCAGGATCAGGAGGAGATTGGAGCCCGTGAGAGAAAATGAGAGGAGATGGGGACCTT 888
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    680 CCCCAGGATCAGGAGGAGATTGGAGCCCGAGAGAAAATGAGAGGAGATGGGGACCTT 739

Qy    889 GACCTGAACGGGCAGTTGTTGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAG 948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    740 GACCTGAACGGGCAGTTGTTGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAG 799

Qy    949 GACGTCAGAGCAGCGCTGGAAGCCAGACAAGATGGAGGAGACCGTCGACATGCATCATC 1008
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    800 GACGTCAGAGCAGCGCTGGGAAGCCAGACAAGATGGAGGAGACCGTCGACATGCATCATC 859

Qy   1009 TGCCAGGACCTGCTGCACGACTGCGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCG 1068
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    860 TGCCAGGACCTGCTGCACGACTGCGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCG 919

Qy   1069 GCTTGCTACTCGGGCTGGATGGAGGGCTCGTCCCGTGTCCCTGTGTCCGTGTCCCGTG 1128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    920 GCTTGCTACTCGGGCTGGATGGAGGGCTCGTCCCGTGTCCCTGTGTCCGTGTCCCGTG 979

Qy   1129 GAGCGGATCTGTAAAAACCACATCCTCAACAACCTCGTGAAGCATACCTCACCAGCAT 1188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    980 GAGCGGATCTGTAAAAACCACATCCTCAACAACCTCGTGAAGCATACCTCACCAGCAT 1039

Qy   1189 CCAGACAAGAGTCGCAGTGAAGAAGATGTGGAAGTATGGATGCCAGGAATAAAATCACT 1248
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1040 CCAGACAAGAGTCGCAGTGAAGAAGATGTGGAAGTATGGATGCCAGGAATAAAATCACT 1099

Qy   1249 CAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTCTGATGAAGAAGGGAGTTCAGAG 1308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1100 CAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTTCTGATGAAGAAGGGAGTTCAGAG 1159

Qy   1309 GACCTGCTGGGAGCTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTC 1368
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1160 GACCTGCTGGGAGCTGTCAGACGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGTC 1219

Qy   1369 GTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGAGGCGGCAGCCTCCCACTGCCCAGCA 1428
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1220 GTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGAGGCGGCAGCCTCCCACTGCCCAGCA 1279

Qy   1429 CCCGAGGGGCGAGCCAGGAGAGCCCCACAGGCCCTGGGGATGCACCCTCCACCCTCCAGC 1488
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1280 CCCGAGGGGCGAGCCAGGAGAGCCCCACAGGCCCTGGGGATGCACCCCACGTTCCGTCAGC 1339

Qy   1489 CTGACGACAGCAGTCCAGGATTACGTGTGCCCCTGCGAAGGAGCCACGCCCCTGTGCACC 1548
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1340 CTGACGACAGCAGTCCAGGATTACGTGTGCCCCTGCGAAGGAGCCACGCCCCTGTGCACC 1399

Qy   1549 TGCTGCTTCCAGCCCATGCGGCGTCTGCGTGCTGAGTACTGGGGTCGCGTGTCGCC 1608
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1400 TGCTGCTTCCAGCCCATGCGGCGTCTGCGTGCTGAGTACTGGGGTCGCGTGTCGCC 1459

Qy   1609 CCTCAGCAGTGTGCGGCGTCTGCAGCCTTTCTGCACCTGTACTGGGGCTGTCACCCGG 1668
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1460 CCTCAGCAGTGTGCGGCGTCTGCAGCCTTTCTGCACCTGTACTGGGGCTGTCACCCGG 1519

Qy   1669 ACCGGCTGCTACGGCTGCCTGGCCCGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTG 1728
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1520 ACCGGCTGCTACGGCTGCCTGGCCCGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTG 1579

Qy   1729 GACGGCGTGCTGAACAACAACAGCTACGAGTCAGAACATCCTGAAGAGATTACCTGGCAACC 1788

Db   2358 CAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTATCTATT 2417
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2569 TTTATTTTAAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGC 2628
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2418 TTTATTTTAAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGC 2477

Qy   2629 ACAAATATATATAATTTATAATTTACAAATTGAC 2661
          ||||||||||||||||||||||||||||||||||
Db   2478 ACAAATATATATAATTTATAATTTACAAATTGAC 2510
```

```
RESULT 10
AK001658
LOCUS       AK001658                3138 bp    mRNA    linear    PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ10796 fis, clone NT2RP4000648, weakly similar
            to TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP0.
ACCESSION   AK001658
VERSION     AK001658.1  GI:7023050
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Aotsuka,S., Yoshikawa,Y.,
            Matsunawa,H., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
            Wakamatsu,A., Nakamura,Y., Nagahari,K., Masuho,Y. and Sasaki,N.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3138)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1..3138
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RP4000648"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2RP4"
                     /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
     CDS             65..1936
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA91817.1"
                     /db_xref="GI:7023051"
                     /translation="MERPEEGKQSPPPQPWGRLLRLGAEBGEPHVLLRKREWTIGRRR
                     GCDLSFPSNKLVSGDHCRIVVDEKSGGVTLEDTSTSGTVINKLKVVKKQTCPLQTGDV
                     IYLVYRKNEPEHVAYLYESLSEKQGWTQESEMVPCCVAQAGLKLLGSSDPPTLASQ
                     SIVITGSSGGGGI5PKGSGPSVASDEVVSFASALPDRKTASFFSSLEPQDQEDLEPEKKK
                     MRGDGDDLDNGQLLVAQPRRNAQTVHBDVRAAAGKPDKMEETITCIICQDLLHDCVSL
                     QPCMHTFCAACYSGWMERSSLCFTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEED
                     VQSMDARNKITQDMLQPKVRRSFSPSDEGGSSEDLLELSDVDSESSDISQPYVVCRQCPE
                     YRRQAAQPPHCPAPEGEPGAPQAQLGDAPPTSVSLTTAVQDYVCPLQGSHALCTCCFQP
                     MPDRRAEREQDPRVAGPQCAVCLQPFCHLYWCTRTGCYGCLAPFCELNLGDKCLLDGV
                     LNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLR
                     SFRELITQYQQNIPASELPVAVTSRPDCYWGRNCRTQVKAHAMKFNHICEQTRFKN"
ORIGIN
```

```
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1
AUTHORS     Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
            Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
            Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
            Masuho,Y.
TITLE       Full-length cDNA sequences
JOURNAL     Patent: EP 1347046-A 2073 24-SEP-2003;
            Research Association for Biotechnology (JP)
FEATURES    Location/Qualifiers
source      1..2448
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
ORIGIN

Query Match           35.2%;  Score 944;  DB 6;  Length 2448;
Best Local Similarity 99.9%;  Pred. No. 0;
Matches 994;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  0;

Qy    708 TGGGTCTGGGGGTGGTGGCATCTCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGA  767
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357 TGGGTCTGGGGGTGGTGGCATCTCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGA  416

Qy    768 AGTCTCCAGCTTTGCCTCAGCTCTCCAGACAGAAAGACTGCGTCCTTTTGTCGTTGGA   827
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    417 AGTCTCCAGCTTTGCCTCAGCTCTCCAGACAGAAAGACTGCGTCCTTTTGTCGTTGGA   476

Qy    828 ACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAAATGAGAGGAGATGGGACCT  887
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    477 ACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAAATGAGAGGAGATGGGACCT  536

Qy    888 TGACCTGAACGGCAGTTGTTGGTCGCACAACCGTGAGAAATGCCAAACCGTCCACGA    947
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    537 TGACCTGAACGGCAGTTGTTGGTCGCACAACCGTGAGAAATGCCAAACCGTCCACGA    596

Qy    948 GGACGTCAGAGCAGCGGCTGGGAAGCCAGACAAGATGGAGAGACGCTGACATGCATCAT 1007
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    597 GGACGTCAGAGCAGCGGCTGGGAAGCCAGACAAGATGGAGAGACGCTGACATGCATCAT  656

Qy   1008 CTGCCAGGACTGCTGCCACGACTGGCTGGAGTGCCGTCCGTCGAGTTGCAGCCGTGTCCGT 1067
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    657 CTGCCAGGACTGCTGCCACGACTGGCTGGAGTGCCGTCCGTCGAGTTGCAGCCGTGTCCGC  716

Qy   1068 GGCTTGCTACTCGGGGTGGATGAGGCCTCGTCCCTGTCCTTTGTCTGTCCGTGTCCGT  1127
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    717 GGCTTGCTACTCGGGGTGGATGAGGCCTCGTCCCTGTCCTTTGTCTGTCCGTGTCCGT  776

Qy   1128 GGAGCGGATCTGTAAAAACCACATCCTCAACAACCTCGTGAAGCATACCTCATCCAGCA 1187
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    777 GGAGCGGATCTGTAAAAACCACATCCTCAACAACCTCGTGAAGCATACCTCATCCAGCA  836

Qy   1188 TCCAGACAAGAGTCGCAGTGAAGATGTGCAAAGTATGGATGCCAAGAATAAAATCAC  1247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    837 TCCAGACAAGAGTCGCAGTGAAGATGTGCAAAGTATGGATGCCAAGAATAAAATCAC  896

Qy   1248 TCAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTTCTGATGAAGAAGGGAGTTCAGA 1307
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    897 TCAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTTCTGATGAAGAAGGGAGTTCAGA  956

Qy   1308 GGACCTGCTGAGCTGTGCAGAGCTTGACAGTTGACAGTCCTCAGACATTAGCCAGCCATCGT 1367
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    957 GGACCTGCTGAGCTGTGCAGAGCTTGACAGTTGACAGTCCTCAGACATTAGCCAGCCATAGT 1016

Qy   1368 CGTGTGCCGGCAGTGTCCAGCAGTGTCCTGAGTACAGAAGGCAGGCGCGCAGCCCACTGCCCAGC 1427
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1017 CGTGTGCCGGCAGTGTCCAGCAGTGTCCTGAGTACAGAAGGCAGGCGCGCAGCCCACTGCCCAGC 1076

Qy   1428 ACCCGAGGGCGAGCCAGGAGCCCACAGGCCCTGGGGGATGCACCCTCCACGTCCGTCAG 1487
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1077 ACCCGAGGGCGAGCCAGGAGCCCACAGGCCCTGGGGGATGCACCCTCCACGTCCGTCAG 1136
```

Lower block (rotated column):

```
Db   1580 GACGGCGGTGCTGAACACAACAGCTACGAGTCAGACATCCTGAAGAATTACCTGGCAACC 1639
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1789 AGAGGTTTGACATGGAAAAACATGTTGACCGGAGAGCCTCGTGGCTCCCAGCGGGGGGAGTG 1848
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1640 AGAGGTTTGACATGGAAAAACATGTTGACCGGAGAGCCTCGTGGCTCCCAGCGGGGGGAGTG 1699

Qy   1849 TTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTTGTTACTGCTGTGGCCTG 1908
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1700 TTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTTGTTACTGCTGTGGCCTG 1759

Qy   1909 CGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCGAGACATTCCTGCTTCCGAGTTGTGCCA 1968
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1760 CGCAGCTTCCGTGAGCTGACCTATCAGTATCAGCGAGAACATTCCTGCTTCCGAGTTGTGCCA 1819

Qy   1969 GTGGCCGTAACATCCCGTCCTGACTCGGGCCGTACTGGGCCGTAACTGCCGCACTCAGGTGAAA 2028
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1820 GTGGCCGTAACATCCCGTCCTGACTCGGGCCGTACTGGGCCGTAACTGCCGCACTCAGGTGAAA 1879

Qy   2029 GCTCACCACGCCATGCCATGAAATTCAATCATATATCTGTGAACAGACAAGGTTCAAAAAACTAAGCA 2088
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1880 GCTCACCACGCCATGATGAAATTCAATCATATATCTGTGAACAGACAAGGTTCAAAAAACTAAGCA 1939

Qy   2089 TCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAGTGAAGAGAGCGTGTTTTTAAAAATACA 2148
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1940 TCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAGTGAAGAGAGCGTGTTTTTAAAAATACA 1999

Qy   2149 GAGACAAGCAAGTCAAGTCAAGGTGTTTTCACAGACCCCCGAGGGAAGGGACGCAGGGTCTCCGA 2208
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2000 GAGGCAAGCAAGTCAAGGTGTTTTCACAGACCCCCGAGGGAAGGGACGCAGGGTCTCCGA 2059

Qy   2209 CAGGTGCTTCTGGGGTGACTCTTCGTGGGAGTCTTTACCCTCGAGTGAGACCCTCCCCA 2268
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2060 CAGGTGCTTCTGGGGTGACTCTTCGTGTGAGC--TTTACCCTCTGAGTGAGACCCTCCCCA 2117

Qy   2269 GAGCCCCGGGGGGCCCAGCCGCGCCCGCCCCTCCTGTGAGCGCTCGGGCAGGGCTCGTCGTGTGTGGCAT 2328
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2118 GAGCCCCGGGGGGCCCAGCCGCGCCCGCCCCTCCTGTGAGCGCTCGGGCAGGGCTCGTCGTGTGTGGCAT 2177

Qy   2329 CAGCAGCAGAGACGAAGCCTTTCTTGTACAAACTACAGTTAAAGCAGAGCTGTGTTTGCAGAGGGGGCAGTTTT 2388
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2178 CAGCAGCAGAGACGAAGCCTTTCTTGTACAAACTACAGTTAAAGCAGAGCTGTGTTTGCAGAGGGGGCAGTTTT 2237

Qy   2389 GCTCTTTGTACATTTGCGAAACTACAGTTAAAGCAGAGTGTGTTTTCAGGAAAAGTT 2448
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2238 GCTCTTTGTACATTTGCGAAACTACAGTTAAAGCAGAGTGTGTTTTCAGGAAAAGTT 2297

Qy   2449 TCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTA 2508
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2298 TCAAGGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTA 2357

Qy   2509 CAGCCTACAGGACGACGTACACAATATCCTGCTCGGGAAAACCACACAGCATTTATCTATTT 2568
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2358 CAGCCTACAGGACGACGTACACAATATCCTGCTCGGGAAAACCACACAGCATTTATCTATTT 2417

Qy   2569 TTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGC 2628
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2418 TTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGC 2477

Qy   2629 ACAAATAATATAATTATAATTTACAAATTGAC 2661
          ||||||||||||||||||||||||||||||||
Db   2478 ACAAATAATAATTATAATTTATAATTTACAAATTGAC 2510
```

```
RESULT 11
AX834949
LOCUS       AX834949    2448 bp    DNA    linear    PAT 15-DEC-2003
DEFINITION  Sequence 2073 from Patent EP1347046.
ACCESSION   AX834949
VERSION     AX834949.1    GI:39921084
KEYWORDS    .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
```

```
Qy   768 AGTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAGACTGCGTCCTTTCGTCGTTGGA 827
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   417 AGTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAGACTGCGTCCTTTCGTCGTTGGA 476

Qy   828 ACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAAATGAGGAGGATGGGGACCT 887
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   477 ACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAAATGAGGAGGATGGGGACCT 536

Qy   888 TGACCTGAACGGGCAGTGTGTTGGTCGCACACCGGTGCACAACCGGTCCACGA 947
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   537 TGACCTGAACGGGCAGTGTGTTGGTCGCACACCGGTGCACAACCGGTCCACGA 596

Qy   948 GGACGTCAGACAGCGGCTGGGAAGCCAGACAAGATGGGAGGAGCACGCTGACATGCATCAT 1007
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   597 GGACGTCAGACAGCGGCTGGGAAGCCAGACAAGATGGGAGGAGCACGCTGACATGCATCAT 656

Qy  1008 CTGCCAGGACCGTGCTGCACGACTGCGTGAGTTGCAGCCGTCCGTGTCCCGT 1067
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   657 CTGCCAGGACCGTGCTGCACGACTGCGTGAGTTGCAGCCGTCCGTGTCCCGT 716

Qy  1068 GGCTTGCTACTCGGGGTGGATGGAGCGGCTCGTCCCTGTGTCGTCCTACTGCGCCGTGTCCCGT 1127
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   717 GGCTTGCTACTCGGGGTGGATGGAGCGGCTCGTCCCTGTGTCGTCCTACTGCGCCGTGTCCCGT 776

Qy  1128 GGAGCGGATCGTAAAAACCACATCCTCAACAACCGTGGAAGCATACCTCATCCAGCA 1187
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   777 GGAGCGGATCGTAAAAACCACATCCTCAACAACCGTGGAAGCATACCTCATCCAGCA 835

Qy  1188 TCCAGACAAAGTGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCAC 1247
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   837 TCCAGACAAAGTGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCAC 896

Qy  1248 TCAAGACATGCTGCAGCCCAAAGTCAGGGCGTCTTTTCTGATGAAGAAGGGAGTTCAGA 1307
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   897 TCAAGACATGCTGCAGCCCAAAGTCAGGGCGTCTTTTCTGATGAAGAAGGGAGTTCAGA 956

Qy  1308 GGACCTGCTGGAGCTGTGACAGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGT 1367
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   957 GGACCTGCTGGAGCTGTGACAGTTGACAGTGAGTCCTCAGACATTAGCCAGCCATACGT 1016

Qy  1368 CGTGTGCCGGCAGTGTCCTGAGTACAGAGGCAGGCGGCGGCAGCCTCCCCACTGCCCAGC 1427
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1017 CGTGTGCCGGCAGTGTCCTGAGTACAGAGGCAGGCGGCGGCAGCCTCCCCACTGCCCAGC 1076

Qy  1428 ACCCGAGGGCAGCAGCCCCACCAGGGCCCTGGGGATGCACCCTCCCAGGTCCTCAG 1487
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1077 ACCCGAGGGCAGCAGCCCCACCAGGGCCCTGGGGATGCACCCTCCCAGGTCCTCAG 1136

Qy  1488 CCTGACGACAGCAGTCCAGGATTACGTGCCCTTCGCAAGGAAGCCACGCCCGTGTGCAC 1547
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1137 CCTGACGACAGCAGTCCAGGATTACGTGCCCTTCGCAAGGAAGCCACGCCCGTGTGCAC 1196

Qy  1548 CTGCTGCTTCCAGCCCATGCCCGACCCGAGCGGAGGCGGAGGCGGAGCGGAGGACCCGCGTCGC 1607
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1197 CTGCTGCTTCCAGCCCATGCCCGACCCGAGCGGAGGCGGAGGCGGAGCGGAGGACCCGCGTCGC 1256

Qy  1608 CCCTCAGCAGTGTGCGGGTGCGGTCTGCCTCAGCCTTCTGCCACCTGTACTGGGGCTGCACCCG 1667
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1257 CCCTCAGCAGTGTGCGGGTGCGGTCTGCCTCAGCCTTCTGCCACCTGTACTGGGGCTGCACCCG 1316

Qy  1668 GACCGGCTGCACGGCTGCACGGCTGCCTGGCCCGTTTGTG 1702
         ||||||||||||||||||||||||||||||||||||||||
Db  1317 GACCGGCTGCACGGCTGCACGGCTGCCTGGCCCGTTTGTG 1351
```

RESULT 12
AK097671

```
LOCUS       AK097671                2448 bp    mRNA            linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ40352 fis, clone TESTI2033505, highly similar
            to Homo sapiens cell cycle checkpoint protein CHFR mRNA.
ACCESSION   AK097671
VERSION     AK097671.1  GI:21757518
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
            Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A., Kawakami,B.,
            Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and
            Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2448)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..2448
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="TESTI2033505"
                     /tissue_type="testis"
                     /clone_lib="TESTI2"
                     /note="cloning vector: pME18SFL3"
ORIGIN
```

Query Match 35.2%; Score 944; DB 9; Length 2448;
Best Local Similarity 99.9%;
Matches 994; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

```
Qy   708 TGGGTCTCGGGGTGGTGGCATCTCCCCTAAAGGAAGTGGTCCCTCTGTTGCAAGTGATGA 767
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357 TGGGTCTCGGGGTGGTGGCATCTCCCCTAAAGGAAGTGGTCCCTCTGTTGCAAGTGATGA 416
```

RESULT 13
BD183291

```
LOCUS       BD183291                6235 bp    DNA             linear   PAT 17-JUN-2003
DEFINITION  Novel genes and proteins encoded by the genes.
ACCESSION   BD183291
VERSION     BD183291.1  GI:31875491
KEYWORDS    JP 2002345492-A/4.
SOURCE      Homo sapiens (human)
```

ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 6235)
AUTHORS   Ohara,O., Nagase,T. and Nakajima,D.
TITLE   Novel genes and proteins encoded by the genes
JOURNAL   Patent: JP 2002345492-A 4 03-DEC-2002;
          KAZUSA DNA RESEARCH INSTITUTE
COMMENT   OS Homo sapiens (human)
          PN JP 2002345492-A/4
          PD 03-DEC-2002
          PF 26-FEB-2002 JP 2002049009
          PI OSAMU OHARA,TAKAHIRO NAGASE,DAISUKE NAKAJIMA
          PC C12N15/09,C07K14/47//A61K31/711,A61K38/00,A61P25/00,
          PC A61P25/14,
          PC A61P25/18,A61P35/00,C12N15/00,A61K37/02
          CC Novel genes and proteins encoded by the genes FH Key
FEATURES   Location/Qualifiers
FT   CDS   (50)..(1414).
source   Location/Qualifiers
                   1..6235
                   /organism="Homo sapiens"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:9606"
ORIGIN

Query Match   34.7%;   Score 930; DB 6; Length 6235;
Best Local Similarity 99.7%;   Pred. No. 0;
Matches 1160; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

Qy   1499 CAGTCCAGGATTACGTGCGCCCCTGCAAGGAGCCACGCCCCTGTGCACCTGCCTTCC 1558
          ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   4447 CAGTCCAGGATTACGTGCGCCCCTCTGCAAGGAGCCACGCCCCTGTGCACCTGCCTTCC 4506

Qy   1559 AGCCCATGCCCGACCGGAGACGGGAGCGGGAGCGCGCAGGACCCGCGTGTCGCCCCTCAGCAGT 1618
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4507 AGCCCATGCCCGACCGGAGACGGGAGCGGGAGCGCGCAGGACCCGCGTGTCGCCCCTCAGCAGT 4566

Qy   1619 GTGCGGTCTGCCTGCAGCCTTTCTGCCACCTGTACTGGGGCTGCACCGGGGCTGCGCT 1678
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4567 GTGCGGTCTGCCTGCAGCCTTTCTGCCACCTGTACTGGGGCTGCACCGGGGCTGCGCT 4626

Qy   1679 ACGGCTGCTGGCCCGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTGGACGGCGTGC 1738
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4627 ACGGCTGCTGGCCCGTTTGTGAGCTCAACCTGGGTGACAAGTGTCTGGACGGCGTGC 4686

Qy   1739 TGAACAACAGCAGTCAGAGTCAGACATCCTGAAGAATTACCTGGCAACCAGAGGTTTGA 1798
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4687 TGAACAACAGCAGTCAGAGTCAGACATCCTGAAGAATTACCTGGCAACCAGAGGTTTGA 4746

Qy   1799 CATGGAAAAACATGTTGACCGAGAGCCTCGTGCTGCCTCCCAGCGGGGGAGTGTTTCTGCTGT 1858
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4747 CATGGAAAAACATGTTGACCGAGAGCCTCGTGCTGCCTCCCAGCGGGGGAGTGTTTCTGCTGT 4806

Qy   1859 CTGATTACAGAGTCACGGGAGACACCGTTCGTGTTACTGCTGTCTGGACGGCCAGCTTCC 1918
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4807 CTGATTACAGAGTCACGGGAGACACCGTTCGTGTTACTGCTGTCTGGACGGCCAGCTTCC 4866

Qy   1919 GTGAGCTGACCTATCAGTATCGGCAGAGACATTCCTGCTTCCAGTTGCCAGTTGGCCGTAA 1978
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4867 GTGAGCTGACCTATCAGTATCGGCAGAGACATTCCTGCTTCCAGTTGCCAGTTGGCCGTAA 4926

Qy   1979 CATCCGTGACTGCTACTGGGGCCGTAACTGCCGCACTGCCCGCACTCAGGTGAAAGTCACCACG 2038
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4927 CATCCGTGACTGCTACTGGGGCCGTAACTGCCGCACTGCCCGCACTCAGGTGAAAGTCACCACG 4986

Qy   2039 CCATGAAATTCAATCATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGCC 2098
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4987 CCATGAAATTCAATCATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGCC 5046

Qy   2099 CTGAGCAGTTTCAGCACTGAGGTGAAGAGACCGTGTTTTAAAATACAGAGACAAGCA 2158
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5047 CTGAGCAGTTTCAGCACTGAGGTGAAGAGACCGTGTTTTAAAATACAGAGACAAGCA 5106

Qy   2159 CGTCAAGGTGTTTCACAGCCCCTGAGGGACGCAGGGACGCAGGGTCTCCGACAGGTGCTCT 2218
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5107 CGTCAAGGTGTTTCACAGCCCCTGAGGGAGGAAGGGACGCAGGGTCTCCGACAGGTGCTCT 5166

Qy   2219 GGGGGTGACTCTTCTGTGGAGCTTTTTACCCTCTGAGTGAGACCCTCCCAGAGCCCCGGG 2278
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db   5167 GGGGGTGACTCTTCTGTGGAGC--TTTACCCTCTGAGTGAGACCCTCCCAGAGCCCCGGG 5224

Qy   2279 GGCCCCAGCCCGCCCTTCCTGTGAGCGCTGGGCAGGCTCGTGGTGGGCATCAGCAGCAGA 2338
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5225 GGCCCCAGCCCGCCCTTCCTGTGAGCGCTGGGCAGGCTCGTGGTGGGCATCAGCAGCAGA 5284

Qy   2339 GACGAGCCTTTCTGTAACATGCGGCCGTCCGCCCAGAGGGGCAGTTTCCTCTTTTGT 2398
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5285 GACGAGCCTTTCTGTAACATGCGGCCGTCCGCCCAGAGGGGCAGTTTTGCTCTTTTGT 5344

Qy   2399 ACATTTTCCGAAACTACAGTAAAGCAGAGAGTCTGTTTCAGGAAAAGTTTCAAGGGAGA 2458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5345 ACATTTTCCGAAACTACAGTAAAGCAGAGAGTCTGTTTCAGGAAAAGTTTCAAGGGAGA 5404

Qy   2459 AGGGCAAGTTATCAAAAACATTGTTCAGGAGAGGGAGACATAGTTACAGCCTACAG 2518
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5405 AGGGCAAGTTATCAAAAACATTGTTCAGGAGAGGGAGACATAAGTTACAGCCTACAG 5464

Qy   2519 GACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTATCTATTTTTATTTTAA 2578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5465 GACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTTTTATTTTAA 5524

Qy   2579 TAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATA 2638
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5525 TAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATA 5584

Qy   2639 TAATTTATAATTTACAAATTGAC 2661
          |||||||||||||||||||||||
Db   5585 TAATTTATAATTTACAAATTGAC 5607

RESULT 14
AK090948
LOCUS   AK090948   2257 bp   mRNA   linear   PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ33629 fis, clone BRAMY2022454, highly similar
            to Homo sapiens cell cycle checkpoint protein CHFR mRNA.
ACCESSION   AK090948
VERSION   AK090948.1 GI:21749208
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
            Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
            Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE   NEDO human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2257)
AUTHORS   Isogai,T. and Yamamoto,J.
TITLE   Direct Submission
JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,

```
         HRI, and Biotechnology Center, National Institute of Technology and
         Evaluation; clone selection for full insert sequencing: HRI and
   RAB; annotation: HRI and RAB.
FEATURES
   source
         1..2257
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="BRAMY2024454"
         /tissue_type="amygdala"
         /clone_lib="BRAMY2"
         /note="cloning vector: pME18SFL3"
ORIGIN

Query Match     33.9%;   Score 909;  DB 9;  Length 2257;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches   959;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;
```

```
Qy   1702 GAGCTCAACCTGGGTGACAAGTGTCTGACGGCGTGCTGAACAACAACAGCTACGAGTCA 1761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669 GAGCTCAACCTGGGTGACAAGTGTCTGACGGCGTGCTGAACAACAACAGCTACGAGTCA 728

Qy   1762 GACATCCGAAGAATTACCTGGCAACCAGAGGTTGACATGGAAAAACATGTTGACCGAG 1821
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729 GACATCCTGAAGAATTACCTGGCAACCAGAGGTTGACATGGAAAAACATGTTGACCGAG 788

Qy   1822 AGCCTCGTGGCCTCCAGCGGGGAGTGTTTCTGCTGTGTCTGATTTACAGAGTCACGGGAGAC 1881
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    789 AGCCTCGTGGCCTCCAGCGGGGAGTGTTTCTGCTGTGTCTGATTTACAGAGTCACGGGAGAC 848

Qy   1882 ACCGTTCTGTGTTACTGCTTCCGTGAGCTGACCTATCAGTAGTCGG 1941
          ||||||||||||||||||||||||||||||||||||||||||
Db    849 ACCGTTCTGTGTTACTGCTTCCGTGAGCTGACCTATCAGTAGTCGG 908

Qy   1942 CAGAACATTCCTGCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCTGACTGCTACTGG 2001
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 CAGAACATTCCTGCTTCCGAGTTGCCAGTGCCGTAACATCCCGTCCTGACTGCTACTGG 968

Qy   2002 GGCCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCCATGAAAATTCAATCATATCTGT 2061
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    969 GGCCGTAACTGCCGCACTCAGGTGAAAGCTCACCACGCCATGAAAATTCAATCATATCTGT 1028

Qy   2062 GAACAGACAAGGTTCAAAAATCAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAG 2121
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1029 GAACAGACAAGGTTCAAAAATCAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAG 1088

Qy   2122 GTGAAGAGAGAGCGTGTTTTAAAATACAGAGACAAGGACACGTCAAGGTGTTTCACAGCCCC 2181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1089 GTGAAGAGAGAGCGTGTTTTAAAATACAGAGACAAGGACACGTCAAGGTGTTTCACAGCCCC 1148

Qy   2182 CTGAGGGAAGGGACGCAGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCTGTGGAGCTT 2241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1149 CTGAGGGAAGGGACGCAGGTCTCCGACAGGTGCTCTGGGGTGACTCTTCTGTGGAGCTT 1208

Qy   2242 TTTACCCTCTGAGTGAGACCCTCCCAGAGCCCGGGGGCCCGCAGCCCCGGGGGCCCTGGTG 2301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1209 TTTACCCTCTGAGTGAGACCCTCCCAGAGCCCGGGGGCCCGCAGCCCCGGGGGCCCTGGTG 1268

Qy   2302 AGCGCTGGGCAGGGCTCATGGTGGCATCAGCAGCAGACGAGAGCCTTTCTGTAACATGC 2361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 AGCGCTGGGCAGGGCTCATGGTGGCATCAGCAGCAGACGAGAGCCTTTCTGTAACATGC 1328

Qy   2362 GGCCGTCCCGCCAGAGAGGGGCAGTTTGCTTTGTTGTACATTTTCCGAAACTACAGTTAA 2421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329 GGCCGTCCCGCCAGAGAGGGGCAGTTTGCTTTGTTGTACATTTTCCGAAACTACAGTTAA 1388

Qy   2422 AGCAGAAGTCTGTTTTCAGGAAAAGTTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATT 2481
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389 AGCAGAAGTCTGTTTTCAGGAAAAGTTTCAAGGGAGAAGGGCAAGTTTATCAAAAACATT 1448

Qy   2482 GTTTCAGGAAGGGAGCATAAGTTTACAGCCTACAGACGTACAACAATATCCTGCTGCT 2541
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1449 GTTTCAGGAAGGGAGCATAAGTTTACAGCCTACAGACGTACAACAATATCCTGCTGCT 1508
```

```
Qy   2542 GGGAAAACCACCAGCATTTATCATCATTTTTATTTTAATTTTAATAGGTTGGTGCTTATCTCTAA 2601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1509 GGGAAAACCACCAGCATTTATCATCATTTTTATTTTAATTTTAATAGGTTGGTGCTTATCTTCTAA 1568

Qy   2602 TAAGATTTAAATGTCACAAACTGTAGCACAAATATATATAATTTATAAATTTACAAATTGAC 2661
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1569 TAAGATTTAAATGTCACAAACTGTAGCACAAATATATATAATTTATAAATTTACAAATTGAC 1628
```

```
RESULT 15
AX713476
LOCUS       AX713476             2186 bp    DNA        linear   PAT 15-APR-2003
DEFINITION  Sequence 160 from Patent EP1293569.
ACCESSION   AX713476
VERSION     AX713476.1  GI:29988332
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
            Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
            Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
            Masuho,Y.
  TITLE     Full-length cDNAs
  JOURNAL   Patent: EP 1293569-A 160 19-MAR-2003;
            Helix Research Institute (JP) ; Research Association for
            Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1..2186
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match     23.6%;   Score 633;  DB 6;  Length 2186;
Best Local Similarity   99.7%;   Pred. No. 0;
Matches   863;  Conservative   0;  Mismatches   1;  Indels   2;  Gaps   1;
```

```
Qy   1796 TGACATGGAAAAACATGTTGACCGAGAGCCTCGTGGCCTCTCCAGCGGGGAGTGTTCTGC 1855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    695 TGACATGGAAAAACATGTTGACCGAGAGCCTCGTGGCCTCTCCAGCGGGGAGTGTTCTGC 754

Qy   1856 TGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTTACTGCTTCGTGGCCTGCGCAGCT 1915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    755 TGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTTACTGCTTCGTGGCCTGCGCAGCT 814

Qy   1916 TCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCG 1975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    815 TCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCG 874

Qy   1976 TAACATCCCGTCCTGACTGCTACTGGGCCGTAACTGCCGCACTCAGGTGAAAGCTCACC 2035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    875 TAACATCCCGTCCTGACTGCTACTGGGCCGTAACTGCCGCACTCAGGTGAAAGCTCACC 934

Qy   2036 ACGCCATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAACTAAGCATCCAGAG 2095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    935 ACGCCATGAAATTCAATCATATCTGTGAACAGACAAGGTTCAAAACTAAGCATCCAGAG 994

Qy   2096 GCCCTGAGCAGCTTTCAGCACTGGAGGTGAGGTGAAGAGAGGCGTGTTTTAAAATACAGAGACAA 2155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    995 GCCCTGAGCAGCTTTCAGCACTGGAGGTGAGGTGAAGAGAGGCGTGTTTTAAAATACAGAGACAA 1054

Qy   2156 GCACGTCAAGGTGTTTCACAGCCCCTGAGGGAAGGGACGCGAGGGCTCTCCGACAGGTGC 2215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1055 ACACGTCAAGGTGTTTCACAGCCCCTGAGGGAAGGGACGCGAGGGCTCTCCGACAGGTGC 1114

Qy   2216 TCTGGGGTGACTCTTCTGTGGAGCTTTTTACCCTCTGAGTGAGACCCTCCCCAGAGCCCC 2275
          |||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db   1115 TCTGGGGTGACTCTTCTGTGGAGC--TTTACCCTCTGAGTGAGACCCTCCCCAGAGCCCC 1172
```

```
Qy  2276  GGGGGCCGCAGCCCGCCCCTCCTGTGAGCGCTGGGCAGGGCTCGTGGTGGCATCAGCAGC  2335
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1173  GGGGGCCCGCAGCCCGCCCCTCCTGTGAGCGCTGGGCAGGGCTCGTGGTGGCATCAGCAGC  1232

Qy  2336  AGAGACGAAGCCTTTCTGTAAACATGCGGCCGTCCCGCCGAGAGGGGCAGTTTTGCTCTTT  2395
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1233  AGAGACGAAGCCTTTCTGTAACATGCGGCCGTCCCGCCGAGAGGGGCAGTTTTGCTCTTT  1292

Qy  2396  TGTACATTTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAAGTTTCAAGGG  2455
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1293  TGTACATTTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAAGTTTCAAGGG  1352

Qy  2456  AGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTA  2515
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1353  AGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTA  1412

Qy  2516  CAGGACGTACACAATATCCTGCTGCTGGGAAAACCACAGCATTTTATCTATTTTTTATTT  2575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1413  CAGGACGTACACAATATCCTGCTCGTCTGGGAAAACCACAGCATTTTATCTATTTTTATTT  1472

Qy  2576  TAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATA  2635
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1473  TAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATA  1532

Qy  2636  ATATAATTTATAATTTACAAATTGAC  2661
          ||||||||||||||||||||||||||
Db  1533  ATATAATTTATAATTTACAAATTGAC  1558

Search completed: May 15, 2004, 05:09:34
Job time : 16432 secs
```

OM nucleic - nucleic search, using sw model

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:         May 15, 2004, 05:09:44 ; Search time 1592 Seconds
                                          (without alignments)
                                          7636.579 Million cell updates/sec

Title:          US-10-048-046-1
Perfect score:  2679
Sequence:       1 aagaattcggcacgaggccg........acaaaaaaaaaaaaaaa 2679

Scoring table:  OLIGO_NUC
                Gapop 60.0 , Gapext 60.0

Searched:       2947324 seqs, 2269024515 residues

Word size :     0

Total number of hits satisfying chosen parameters:    5894648

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :      Published_Applications_NA:*
   1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
   2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
   3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
   4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
   5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
   6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
   7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
   8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
   9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
  10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
  11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
  12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
  13:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
  14:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
  15:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
  16:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
  17:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
  18:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
  19:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1995	74.5	1995	9	US-09-780-525-1	Sequence 1, Appli
2	944	35.2	2448	16	US-10-108-260A-2073	Sequence 2073, Ap
3	765	28.6	1311	9	US-09-764-864-22	Sequence 22, Appl
4	633	23.6	2186	16	US-10-094-749-160	Sequence 160, App
5	455	17.0	693	9	US-09-764-864-483	Sequence 483, App
6	267	10.0	357	9	US-09-764-864-322	Sequence 322, App
7	238	8.9	354	9	US-09-764-864-323	Sequence 323, App
8	238	8.9	354	9	US-09-764-864-742	Sequence 742, App
9	217	8.1	449	10	US-09-918-995-13662	Sequence 13662, A
10	77	2.9	476	10	US-09-918-995-2180	Sequence 2180, Ap
11	60	2.2	60	10	US-09-908-975-6220	Sequence 6220, Ap
12	25	0.9	1085	15	US-10-304-928-15	Sequence 15, Appl
13	25	0.9	1963	15	US-09-957-664-1	Sequence 1, Appli
14	25	0.9	1963	15	US-10-154-419-15	Sequence 15, Appl
c 15	25	0.9	2000	9	US-09-938-842A-3337	Sequence 3337, Ap
c 16	25	0.9	2000	11	US-09-938-842A-3337	Sequence 3337, Ap
17	24	0.9	199	9	US-09-867-701-9622	Sequence 9622, Ap
18	24	0.9	222	9	US-09-764-846-50	Sequence 50, Appl
19	24	0.9	222	9	US-09-764-846-122	Sequence 122, App
20	24	0.9	222	15	US-10-091-483-50	Sequence 50, Appl
21	24	0.9	222	15	US-10-091-483-122	Sequence 122, App
22	24	0.9	320	13	US-10-085-783A-21114	Sequence 21114, A
23	24	0.9	320	16	US-10-242-535A-21114	Sequence 21114, A
24	24	0.9	474	10	US-09-918-995-10672	Sequence 10672, A
c 25	24	0.9	635	13	US-10-027-632-147186	Sequence 147186,
c 26	24	0.9	635	16	US-10-027-632-147186	Sequence 147186,
27	24	0.9	844	17	US-10-444-795B-796	Sequence 796, App
28	24	0.9	11869	9	US-09-764-869-2292	Sequence 2292, Ap
29	24	0.9	11869	15	US-10-091-504-2292	Sequence 2292, Ap
30	24	0.9	11869	16	US-10-227-577-2292	Sequence 2292, Ap
c 31	24	0.9	358246	16	US-10-292-798-1095	Sequence 1095, Ap
32	23	0.9	160	13	US-10-085-783A-54104	Sequence 54104, A
33	23	0.9	160	16	US-10-242-535A-54104	Sequence 54104, A
34	23	0.9	222	9	US-09-867-701-9978	Sequence 9978, Ap
35	23	0.9	242	10	US-09-764-891-2578	Sequence 2578, Ap
36	23	0.9	292	13	US-10-424-599-142624	Sequence 142624,
37	23	0.9	306	13	US-10-424-599-66738	Sequence 66738, A
38	23	0.9	321	13	US-10-424-599-52665	Sequence 52665, A
39	23	0.9	330	9	US-09-783-590-4652	Sequence 4652, Ap
c 40	23	0.9	422	9	US-09-960-352-3608	Sequence 3608, Ap
41	23	0.9	469	10	US-09-918-995-13017	Sequence 13017, A
c 42	23	0.9	470	9	US-09-770-444-201	Sequence 201, App
43	23	0.9	493	10	US-09-918-995-19565	Sequence 19565, A
44	23	0.9	555	13	US-10-424-599-129887	Sequence 129887,
45	23	0.9	687	13	US-10-424-599-115631	Sequence 115631,

ALIGNMENTS

```
RESULT 1
US-09-780-525-1
; Sequence 1, Application US/09780525
; Patent No. US2002000422 3A1
; GENERAL INFORMATION:
; APPLICANT: Bin-Bing Zhou
; APPLICANT: Yuan Zhu
; APPLICANT: Priya Chaturvedi
; APPLICANT: Mark R. Hurle
; APPLICANT: Xiaotong Li
; TITLE OF INVENTION: FHAR1, A NEW RING FINGER PROTEIN
; FILE REFERENCE: GP-70668-C1
; CURRENT APPLICATION NUMBER: US/09/780,525
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 09/456,876
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1995
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-780-525-1

Query Match             74.5%;  Score 1995;  DB 9;  Length 1995;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1995;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy  91 ATGGACGCGCCCGAGGAGGAGGCAAGCAGTCGCGCGCCGCAGCCTCGGGGACCGGCTCCTG 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGACGCGCCCGAGGAGGAGGCAAGCAGTCGCGCGCCGCAGCCTCGGGGACCGGCTCCTG 60

Qy 151 CGTCTGGGCGCGGAGGAGGAGGGGCGAGCCCGACCCACAGTCCTCCTGAGGAAGCGGGAGTGGACCATC 210
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 CGTCTGGGCGCGGAGGAGGAGGGGCGAGCCCGACCCACAGTCCTCCTGAGGAAGCGGGAGTGGACCATC 120

Qy 211 GGGCGGAGACGAGGTTGCGACCTTTCCTTCCCCAGCAATAAACTGGTCTCTGGAGATCAC 270
```

```
Db   1201 GAAGAAGGGAGTTCAGAGGACCTGCTGGAGACGTGTCAGACGTTGACAGTGAGTCCTCAGAC 1260
                                                                          
Qy   1351 ATTAGCCAGCCATACGTCGTGTCCGCAGTGTCCTGAGTACGAGTACAGAGGCGGCGGCGCAG 1410
          |||||||||||||||||||||||||||||||||||||||    |||||||||||||||||||
Db   1261 ATTAGCCAGCCATACGTCGTGTCCGCAGTGTCCTGAGTACAGAGAGGCAGGCGGCGGCGGAG 1320

Qy   1411 CCTCCCACTGCCCAGCACCCGAGGGGCGAGCCAGGAGCCCACAGCCCTGGGGGATGCA 1470
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CCTCCCACTGCCCAGCACCCGAGGGGCGAGCCAGGAGCCCACAGCCCTGGGGGATGCA 1380

Qy   1471 CCCTCCAGTCCGTCAGCCTGAGCGACGACAGCAGTCCAGGATTACGTGTGCCCTCTGCAAGGA 1530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CCCTCCAGTCCGTCAGCCTGAGCGACGACAGCAGTCCAGGATTACGTGTGCCCTCTGCAAGGA 1440

Qy   1531 AGCCACGCCCTGTCGCACCTGCTTCCAGCCAGTGTGCGGTCGTGCCTGCGACAGTGTGCCACCTG 1590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AGCCACGCCCTGTCGCACCTGCTTCCAGCCAGTGTGCGGTCGTGCCTGCGACAGTGTGCCACCTG 1500

Qy   1591 CAGGACCGGCGTGTGCCACCGGGACCGGCTGCTACGGCTGCCCCGTTTTGTGAGCTCAAC 1650
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CAGGACCGGCGTGTGCCACCGGGACCGGCTGCTACGGCTGCCCCGTTTTGTGAGCTCAAC 1560

Qy   1651 TACTGGGGCTGCAACGGCGTGCTGACGGCGTGCTGAACAACAGCTAGGAGTCAGACATCCTG 1710
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TACTGGGGCTGCAACGGCGTGCTGACGGCGTGCTGAACAACAGCTAGCAGTCAGACATCCTG 1620

Qy   1711 CTGGGTGACAAGTGTCTGCACCACGAGGTTGACATGGAAAAACATGTTGACCGAGAGCCTCGTG 1770
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 CTGGGTGACAAGTGTCTGCACCACCAGAGGTTGACATGGAAAAACATGTTGACCGAGAGCCTCGTG 1830

Qy   1771 AAGAATTACCTGGCAACCAGAGGTTTCGTGCTTGTCTGATTACAGAGTCACGGGAGACACCGTTCTG 1890
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 AAGAATTACCTGGCAACCAGAGGTTTCGTGCTTGTCTGATTACAGAGTCACGGGAGACACCGTTCTG 1740

Qy   1831 GCTCTCCAGCGGGGAGTGTTTCTGCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATT 1950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GCTCTCCAGCGGGGAGTGTTTCTGCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATT 1860

Qy   1891 TGTTACTGCTGTGGCCTGCCAGTGTGCCGCAGCTTCCGTAACACTCCCGTCCTACTGGGCCGTAAC 2010
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGTTACTGCTGTGGCCTGCCAGTGTGCCGCAGCTTCCGTAACACTCCCGTCCTACTGGGCCGTAAC 1920

Qy   1951 CCTGCTTCCGAGTTGCCAGTTGCCAGTGCCGTAACATCCCGTCCTACTGGGCCCGTAAC 2070
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CCTGCTTCCGAGTTGCCAGTTGCCAGTGCCGTAACATCCCGTCCTACTGGGCCCGTAAC 1980

Qy   2011 TGCCGCACTCAGGTGAAAGTCCACCACGCCATGAAATTCAATCATATCTGTGAACAGACA 2070
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TGCCGCACTCAGGTGAAAGTCCACCACGCCATGAAATTCAATCATATCTGTGAACAGACA 1980

Qy   2071 AGTTCAAAAACTAA 2085
          ||||||||||||||
Db   1981 AGTTCAAAAACTAA 1995
```

```
Db    121 GGGCGGACGAGAGTTGCGACCTTCCTTCCCAGCAATAAACTGGTCTCTGGAGATCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    271 TGTAGAATTGTAGTGGATGAAAAATCAGGTCAGGTGACACTGAAGATACCAGCACCAGT 330
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGTAGAATTGTAGTGGATGAAAAATCAGGTCAGGTGACACTGAAGATACCAGCACCAGT 240

Qy    331 GGAACAGTGATTAACAGCTGAAGGTTGTTAAGAGCAGACATGCCCTTTACAGACTGGG 390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGAACAGTGATTAACAGCTGAAGGTTGTTAAGAGCAGACATGCCCTTTACAGACTGGG 300

Db    391 GATGTCATCTACTTGGTTGGTACAGGAAGAATGAACCGGAACACAACGTGGCATACCTCTAT 450
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    301 GATGTCATCTACTTGGTTGGTACAGGAAGAATGAACCGGAACACAACGTGGCATACCTCTAT 360

Db    451 GAATCTTTAAGTGAAAAGCAAGGCATGACACAAGAATCCTTTGAAGCTAACAAGGAAAAT 510
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    361 GAATCTTTAAGTGAAAAGCAAGGCATGACACAAGAATCCTTTGAAGCTAACAAGGAAAAT 420

Db    511 GTGTTCCATGGGACCAAAGATACCTCAGTGCAGGTGCAGTGTGCTTTGAGGACCACAACATCG 570
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    421 GTGTTCCATGGGACCAAAGATACCTCAGTGCAGGTGCAGTGTGCTTTGAGGACCACAACATCG 480

Db    571 GTCCCTCCGTCGTCGCCCACTCAGGTCTTCTCCACGGACGCCTTCTCCTGCGAGGGCGA 630
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    481 GTCCCTCCGTCGTCGCCCACTCAGGTCTTCTCCACGGACGCCTTCTCCTGCGAGGGCGA 540

Db    631 ACGTCAGACCTCTTCCCCACAGCCTCGGCCTCGGGGTGGGGGTGCATCTCCCCCTAAAGGAGGTGGTCCC 690
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    541 ACGTCAGACCTCTTCCCCACAGCCTCGGCCTCGGGGTGGGGGTGCATCTCCCCCTAAAGGAGGTGGTCCC 600

Db    691 GAGCGTTCCTCCAGTTGTGGGTCTGGGGTGTGGGGGTGGGGGTGGTGCATCTCCCCCTAAAGGAGGTGGTCCC 750
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601 GAGCGTTCCTCCAGTTGTGGGTCTGGGGTGTGGGGGTGGGGGTGGTGCATCTCCCCCTAAAGGAGGTGGTCCC 660

Db    751 TCTGTGCAAGTGATGATGATGAAGTCTCCAGCTTTGCCTCAGCTTTGCCTCAGCTCTCCAGAGGATCAGGAGAAAGACTGCG 810
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    661 TCTGTGCAAGTGATGATGATGAAGTCTCCAGCTTTGCCTCAGCTTTGCCTCAGCTCTCCAGAGGATCAGGAGAAAGACTGCG 720

Db    811 TCCTTTTCGTCGTTGGAAGCTCTCCAGCTTTGCCTCAGCTTTGCCTCAGCTCTCCAGGATTTGGAGCCCGTGAAGAAGAAATG 870
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    721 TCCTTTTCGTCGTTGGAAGCTCTCCAGCTTTGCCTCAGCTTTGCCTCAGCTCTCCAGGATTTGGAGCCCGTGAAGAAGAAATG 780

Db    871 AGAGGAGATGGGGACCTTGACCTGAACGGGCAGTTGTTGGTCGCACAACCGCCGTAGAAAT 930
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    781 AGAGGAGATGGGGACCTTGACCTGAACGGGCAGTTGTTGGTCGCACAACCGCCGTAGAAAT 840

Db    931 GCCCAAACCGTCCACGAGGACGTCAGAGGACGCGCCAGCAGCGCCAGCAAGATGGGAGGAG 990
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    841 GCCCAAACCGTCCACGAGGACGTCAGAGGACGCGCCAGCAGCGCCAGCAAGATGGGAGGAG 900

Db    991 ACGCTGACATGCATCATCTGCCAGGACCTTGCTACTCGGGCGCTGGATGGAGCGCTCGTGTCCTGTGTCCT 1050
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    901 ACGCTGACATGCATCATCTGCCAGGACCTTGCTACTCGGGCGCTGGATGGAGCGCTCGTGTCCTGTGTCCT 960

Db   1051 ATGCACACGTTCTGCGCGGCTTGCTACTCGGGCGCTGGATGGAGCGCTCGTGTCCTGTGTCCT 1110
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    961 ATGCACACGTTCTGCGCGGCTTGCTACTCGGGCGCTGGATGGAGCGCTCGTGTCCTGTGTCCT 1020

Db   1111 ACCTGCCGCGTCTGTCCCGTGCCGGATCTGTAAAAAACCACATCCTCAAACAACCTCGTGAA 1170
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1021 ACCTGCCGCGTCTGTCCCGTGCCGGATCTGTAAAAAACCACATCCTCAAACAACCTCGTGAA 1080

Db   1171 GCATACCTCATCCAGCATCCAGACAGAGACTCGCAGTGAAGATGTGCAAAGTATGGAT 1230
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1081 GCATACCTCATCCAGCATCCAGACAGAGACTCGCAGTGAAGATGTGCAAAGTATGGAT 1140

Db   1231 GCCAGGAATAAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTTTCTGAT 1290
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1141 GCCAGGAATAAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGCGGTCTTTTTTCTGAT 1200

Qy   1291 GAAGAAGGGAGTTCAGAGGACCTGCTGAGCTGTCGAGCAGTGAGTCGAGTCCTCAGAC 1350
```

Left alignment block

```
Query Match         35.2%;   Score 944;  DB 16;  Length 2448;
Best Local Similarity 99.9%;   Pred No. 0;
Matches   994;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    0;

Qy   708  TGGGGTCGGGGGTGGTGGCATCTCCCCTAAAGGAGGAGTGGTCCCCTGTCGGCAAGTGATGA  767
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357  TGGGGTCGGGGGTGGTGGCATCTCCCCTAAAGGAGGAGTGGTCCCCTGTCGGCAAGTGATGA  416

Qy   768  AGTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAGACTGCGTCCTTTTCGTCGTTGGA  827
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   417  AGTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAGACTGCGTCCTTTTCGTCGTTGGA  476

Qy   828  ACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAATGAGGAGGATGGGGACCT  887
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   477  ACCCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAATGAGGAGGATGGGGACCT  536

Qy   888  TGACCTGAACGGGCAGTTGTTGGTGCACAACCGGTAGAAATGCCCAAACCGTCCACGA  947
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   537  TGACCTGAACGGGCAGTTGTTGGTGCACAACCGGTAGAAATGCCCAAACCGTCCACGA  596

Qy   948  GGAGCTCAGAGCAGCGGCTGGGAAGCCAGACAAGAGGAGACGCTGACATGCATCAT  1007
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   597  GGACGTCAGAGCAGCGGCTGGGAAGCCAGACAAGAGGAGACGCTGACATGCATCAT  656

Qy  1008  CTGCCAGGACCTGCTGCAGACTGCGTGAGTTTGCAGCCCTGCATGCATGCACGCTCTGCGGC  1067
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   657  CTGCCAGGACCTGCTGCAGACTGCGTGAGTTTGCAGCCCTGCATGCATGCACGCTCTGCGGC  716

Qy  1068  GGCTTGCTACTCGGGCTGGATGGAGGAGCGCTCGTCCCGTGTCCGCGTGTCCCGT  1127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   717  GGCTTGCTACTCGGGCTGGATGGAGGAGCGCTCGTCCCGTGTCCGCGTGTCCCGT  776

Qy  1128  GGAGCGGATCGTAAAAACCACATCTCAACAACCTCGTGGAAGCATACCTCATCCAGCA  1187
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   777  GGAGCGGATCGTAAAAACCACATCTCAACAACCTCGTGGAAGCATACCTCATCCAGCA  836

Qy  1188  TCCAGACAAGAGTCGCAGTGAGAAGATGTGCAAAGTCCAGGAATAAAATCAC  1247
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   837  TCCAGACAAGAGTCGCAGTGAGAAGATGTGCAAAGTCCAGGAATAAAATCAC  896

Qy  1248  TCAAGACATGCTGCGAGCCCAAAGTCAGGCGGTCTCTTTTTCTGATGAAGAAGAGGAGTTCAGA  1307
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   897  TCAAGACATGCTGCGAGCCCAAAGTCAGGCGGTCTCTTTTTCTGATGAAGAAGAGGAGTTCAGA  956

Qy  1308  GGACCTGCTCGGAGCTGTGTCAGACGTTGACAGTGAGTCCTCAGAACATTAGCCAGCCATACGT  1367
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   957  GGACCTGCTCGGAGCTGTGTCAGACGTTGACAGTGAGTCCTCAGAACATTAGCCAGCCATACGT  1016

Qy  1368  CGTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCGGCAGCCCTCCCCACTGCCCAGC  1427
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1017  CGTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCGGCAGCCCTCCCCACTGCCCAGC  1076

Qy  1428  ACCCGAGGGGCGAGCCAGGAGCCCAGAGCCCCAGGGGATGCACCCTCCGTCCGTGTCAG  1487
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1077  ACCCGAGGGGCGAGCCAGGAGCCCAGAGCCCCAGGGGATGCACCCTCCGTCCGTGTCAG  1136

Qy  1488  CCTGACGACAGCAGTCCAGGATTACGTGTGCCCCTGCAAGGAAGCCACGCCCTGTGCAC  1547
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1137  CCTGACGACAGCAGTCCAGGATTACGTGTGCCCCTGCAAGGAAGCCACGCCCTGTGCAC  1196

Qy  1548  CTGCTGCTTCCAGCCCATGCCCGACCGGAGCCGGAGCGGCAGGACCCGGCGGTGTCGC  1607
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1197  CTGCTGCTTCCAGCCCATGCCCGACCGGAGCCGGAGCGGCAGGACCCGGCGGTGTCGC  1256

Qy  1608  CCCTCAGCAGTGTGCGGCTCGGTGCCACTCTGCACCCTGTACTGGGGCTGCACCCG  1667
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1257  CCCTCAGCAGTGTGCGGCTCGGTGCCACTCTGCACCCTGTACTGGGGCTGCACCCG  1316

Qy  1668  GACCCGGCTGCTACGGCTGCCTGGCCCGTTTTGTG  1702
          ||||||||||||||||||||||||||||||||||
Db  1317  GACCCGGCTGCTACGGCTGCCTGGCCCGTTTTGTG  1351
```

Right alignment block

```
RESULT 3
US-09-764-864-22
; Sequence 22, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1311
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1281)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-22

Query Match         28.6%;   Score 765;  DB 9;  Length 1311;
Best Local Similarity 100.0%;   Pred. No. 0;
Matches   765;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   710  GGTCTGGGGGGTGGTGGCATCTCCCCTAAAGGAGGAGTGGTCCCCTGTCGGCAAGTGATGATGAAG  769
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   383  GGTCTGGGGGGTGGTGGCATCTCCCCTAAAGGAGGAGTGGTCCCCTGTCGGCAAGTGATGATGAAG  442

Qy   770  TCTCCAGCTTGCCTCAGCTCTCCCAGACAGAAGACTGCGTCCTTTTCGTCGTTGGAAC  829
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   443  TCTCCAGCTTGCCTCAGCTCTCCCAGACAGAAGACTGCGTCCTTTTCGTCGTTGGAAC  502

Qy   830  CCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAATGAGGAGGATGGGGACCTTG  889
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   503  CCCAGGATCAGGAGGATTTGGAGCCCGTGAAGAAGAAATGAGGAGGATGGGGACCTTG  562

Qy   890  ACCTGAACGGGCAGTGTTGTTGGTCCACAACCGGTAGAAATGCCCAAACCGTCCACGAGG  949
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   563  ACCTGAACGGGCAGTGTTGTTGGTGCACAACCGGTAGAAATGCCCAAACCGTCCACGAGG  622

Qy   950  ACGTCAGAGCAGCGGCTGGGAAGCCAGACAAGATGGAGACAAGATGGAGACGCTGACATGCATCATCT  1009
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   623  ACGTCAGAGCAGCGGCTGGGAAGCCAGACAAGATGGAGACAAGATGGAGACGCTGACATGCATCATCT  682

Qy  1010  GCCAGGACCTGCTGCAGACTGCGTGAGTTTGCAGCCCTGTCATGCACACGTTCTGCGCGG  1069
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   683  GCCAGGACCTGCTGCAGACTGCGTGAGTTTGCAGCCCTGTCATGCACACGTTCTGCGCGG  742

Qy  1070  CTTGCTACTCGGGCTGATGGAGGACGCTCGTCCCGTGTCCGTGTCCGTGTCCGTGG  1129
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   743  CTTGCTACTCGGGCTGATGGAGGACGCTCGTCCCGTGTCCGTGTCCGTGTCCGTGG  802

Qy  1130  AGCGGGATCGTGTAAAAACCACATCTCAACAACCTCGTGGAAGCATACCTCATCCAGCATC  1189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   803  AGCGGGATCGTGTAAAAACCACATCTCAACAACCTCGTGGAAGCATACCTCATCCAGCATC  862

Qy  1190  CAGACAAGAGTCGCAGTGAGAAGAATGTGCAAAGTCGGCCGGTCTTTTTCTTCACTC  1249
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   863  CAGACAAGAGTCGCAGTGAGAAGAATGTGCAAAGTCGGCCGGTCTTTTTCTTCACTC  922

Qy  1250  AAGACATGCTGCAGCCCAAAGTCAGGCCGTCTTTTTCTGATGAAGAAAGGGAGTTCAGAGG  1309
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   923  AAGACATGCTGCAGCCCAAAGTCAGGCCGTCTTTTTCTGATGAAGAAAGGGAGTTCAGAGG  982

Qy  1310  ACCTGCTGGAGCGTGCAGACGTTGACAGTGAGTCCTCAGACACTTAGCCAGCCATACGTCG  1369
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   983  ACCTGCTGGAGCGTGCAGACGTTGACAGTGAGTCCTCAGACACTTAGCCAGCCATACGTCG  1042

Qy  1370  TGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGCGGCGGCAGCCCTCCCCACTGCCCAGCAC  1429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   2156   GCACGTCAAGGTGTTTCACAGCCCCTGAGGAAGGACGGCAGGGTCTCCGACAGGTGC  2215
            |||||||||||||||||||||||||||||| |||| ||||||||||||||||||||||
Db   1055   ACACGTCAAGGTGTTTTCACAGCCCCCTGAGGGAGGGACGGCAGGGTCTCCGACAGGTGC  1114

Qy   2216   TCTGGGGTGACTCTTCTGTGGAGCTTTTACCCTCTGAGTGAGACCCTCCCCAGAGCCCC  2275
            |||||||||||||||||||| |||  ||||||||||||||||||||||||||||||||||
Db   1115   TCTGGGGTGACTCTTCTGTGTGAGC--TTTACCCTCTGAGTGAGACCCTCCCCAGAGCCCC  1172

Qy   2276   GGGGGCCCCAGCCCGCCCCTCCGTGTGAGCGCTCGGCGCTGTGGTGGCATCAGCAGC  2335
            ||||||  ||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   1173   GGGGGCGCGCAGCCCGCCCCTCCTGGTGAGCGCTCGGCGCTGTGGTGGCATCAGCAGC  1232

Qy   2336   AGAGACGAAGCCTTTCGTAACATGGCGGCCGTCCCGCCGAGGAGGGCAGTTTGCTCTTT  2395
            |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   1233   AGAGACGAAGCCTTTCGTAACATGGCGGCCGTCCCGCCGAGGAGGGCAGTTTTGCTCTTT  1292

Qy   2396   TGTACATTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAGTTTCAAGGG  2455
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1293   TGTACATTTCCGAAACTACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAGTTTCAAGGG  1352

Qy   2456   AGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTA  2515
            ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1353   AGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTA  1412

Qy   2516   CAGGACGTACACAATACCTCGCTGCTGGGAAAACCACCAGCATTTATCTATTTTTTATTT  2575
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1413   CAGGACGTACACAATATCCTCGCTGCTGGGAAAACCACCAGCATTTATCTATTTTTTATTT  1472

Qy   2576   TAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATA  2635
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1473   TAATAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATA  1532

Qy   2636   ATATAATTTATAATTTACAAATTGAC  2661
            ||||||||||||||||||||||||||
Db   1533   ATATAATTTATAATTTACAAATTGAC  1558
```

RESULT 5
US-09-764-864-483

```
;   Sequence 483, Application US/09764864
;   Patent No. US20020132753A1
;   GENERAL INFORMATION:
;     APPLICANT: Rosen et al.
;     TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;     FILE REFERENCE: PTZ23
;     CURRENT APPLICATION NUMBER: US/09/764,864
;     CURRENT FILING DATE: 2001-01-17
;     Prior application data removed - consult PALM or file wrapper
;     NUMBER OF SEQ ID NOS: 1792
;     SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 483
;     LENGTH: 693
;     TYPE: DNA
;     ORGANISM: Homo sapiens
;     FEATURE:
;     NAME/KEY: SITE
;     LOCATION: (11)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: SITE
;     LOCATION: (13)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: SITE
;     LOCATION: (623)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: SITE
;     LOCATION: (640)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: SITE
;     LOCATION: (670)
;     OTHER INFORMATION: n equals a,t,g, or c
;     NAME/KEY: SITE
;     LOCATION: (687)
```

```
Db   1043   TGTGCCGCAGTGTCCTGAGTACAGAGAGGCGGGCGCCTCCCACTGCCAGCAC  1102
Qy   1430   CCGAGGGCGAGCCAGGAGCCCCACAGCCCTGGGGATGCACCCT  1474
            |||||||||||||||||||||||||| ||||||||||||||||
Db   1103   CCGAGGGCGAGCCAGGAGCCCCACAGCCCCTGGGGATGCACCCT  1147
```

RESULT 4
US-10-094-749-160

```
;   Sequence 160, Application US/10094749
;   Publication No. US20030219741A1
;   GENERAL INFORMATION:
;     APPLICANT: ISOGAI, TAKAO
;     APPLICANT: SUGIYAMA, TOMOYASU
;     APPLICANT: OTSUKI, TETSUJI
;     APPLICANT: WAKAMATSU, AI
;     APPLICANT: SATO, HIROYUKI
;     APPLICANT: ISHII, SHIZUKO
;     APPLICANT: YAMAMOTO, JUN-ICHI
;     APPLICANT: ISONO, YUUKO
;     APPLICANT: HIO, YURI
;     APPLICANT: OTSUKA, KAORU
;     APPLICANT: NAGAI, KEIICHI
;     APPLICANT: IRIE, RYOTARO
;     APPLICANT: TAMECHIKA, ICHIRO
;     APPLICANT: SEKI, NAOHIKO
;     APPLICANT: YOSHIKAWA, TSUTOMU
;     APPLICANT: OTSUKA, MOTOYUKI
;     APPLICANT: NAGAHARI, KENJI
;     APPLICANT: MASUHO, YASUHIKO
;     TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
;     FILE REFERENCE: 084335/0160
;   CURRENT APPLICATION NUMBER: US/10/094,749
;     CURRENT FILING DATE: 2002-03-12
;     PRIOR APPLICATION NUMBER: 60/350,435
;     PRIOR FILING DATE: 2002-01-24
;     PRIOR APPLICATION NUMBER: JP 2001-328381
;     PRIOR FILING DATE: 2001-09-14
;     NUMBER OF SEQ ID NOS: 3381
;     SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 160
;     LENGTH: 2186
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-10-094-749-160
```

```
Query Match           23.6%;  Score 633;  DB 16;  Length 2186;
Best Local Similarity 99.7%;  Pred. No. 2e-308;
Matches  863;  Conservative  0;  Mismatches  1;  Indels  2;  Gaps  1;
```

```
Qy   1796   TGACATGGAAAAACATGTTGACCGAGAGCCTCGTGGCTCTCCAGCGGCGGGAGTGTTTCTGC  1855
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    695   TGACATGGAAAAACATGTTGACCGAGAGCCTCGTGGCTCTCCAGCGGCGGGAGTGTTTCTGC   754

Qy   1856   TGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGTGGCCTGCCCAGCT  1915
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    755   TGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGTGGCCTGCCCAGCT   814

Qy   1916   TCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCG  1975
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    815   TCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCG   874

Qy   1976   TAACATCCGTCCGACTGCTACTGGGCCGTAACTGCCGACTCAGGTGAAAGTCACC  2035
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    875   TAACATCCGTCCGACTGCTACTGGGCCGTAACTGCCGACTCAGGTGAAAGTCACC   934

Qy   2036   ACGCCATGAAATTCAATCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAG  2095
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    935   ACGCCATGAAATTCAATCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAG   994

Qy   2096   GCCCTGAGCGAGCTTCAGCACTGAGGTGAAGAGAGAGCGGCCGTGTTTTTAAAATACAGAGACAA  2155
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    995   GCCCTGAGCGAGCTTCAGCACTGAGGTGAAGAGAGAGCGGCCGTGTTTTTAAAATACAGAGACAA  1054
```

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (690)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-483

Query Match    17.0%;  Score 455;  DB 9;  Length 693;
Best Local Similarity  100.0%;  Pred. No. 1.3e-218;
Matches  455;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy 1044 GCCCTGCGATGCACACGTTCGCGCGGCTTGCTACTCGGCGGCTGGATGGAGGGCTCGTCCCT 1103
         ||||||||||||||||||||  ||||||||||||||||||  |||||||||| ||||||||||
Db   14 GCCCTGCATGCACACGTTCTGCGCGGCTTGCTACTCGGCGTGGATGGAGCGCTCGTCCCT 73

Qy 1104 GTGTCCTACCTGCGCGTCTCCCGTGAGCGGATCGTGAAAAAACCACCATCCTCAACAACCT 1163
         ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   74 GTGTCCTACCTGCGCGTCCCCGTGAGCGGATCGTGAAAAAACCACCATCCTCAACAACCT 133

Qy 1164 CGTGAAGCATACCTCATCCAGCATCCAGACAACAAGACTCGCAGTGAAGAGATGTGCAAAG 1223
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  134 CGTGGAAGCATACCTCATCCAGCATCCAGACAACAAGAGTCGCAGTGAAGAGATGTGCAAAG 193

Qy 1224 TATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGCCGGTCTTT 1283
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  194 TATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGGCCGGTCTTT 253

Qy 1284 TTCTGATGAAGAGGAGTTCAGAGGACCTGCTCGGAGCGTGTCACAGCGTTGACAGTGAGTC 1343
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  254 TTCTGATGAAGAGGAGTTCAGAGACCTGCTCGGAGCGTCGTCAGACGTTGACAGTGAGTC 313

Qy 1344 CTCAGACATTAGCCAGCCATACGTCGTGTCGCCGCAGTGCCTGAGTACAGAAGGCAGGC 1403
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  314 CTCAGACATTAGCCAGCCATACGTCGTGTCGCCGCAGTGTCCTGAGTACAGAAGGCAGGC 373

Qy 1404 GGGCAGCCTCCCCACTGCCCAGCACCCGAGGGCGAGCCAGGAGGCCCCACAGGCCCTGGG 1463
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  374 GGGCAGCCTCCCCACTGCCCAGCACCCGAGGGCGAGCCAGGAGGCCCCACAGGCCCTGGG 433

Qy 1464 GGATGCACCCTCCAGTCCGTCAGCCTGACGACAG 1498
         |||||||||||||| |||||||||||||||||||
Db  434 GGATGCACCCTCCACGTCCGTCAGCCTGACGACAG 468


RESULT 6
US-09-764-864-322
; Sequence 322, Application US/09764864
; Patent No. US2002013753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 322
; LENGTH: 357
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (300)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-322

Query Match    10.0%;  Score 267;  DB 9;  Length 357;
Best Local Similarity  100.0%;  Pred. No. 9.9e-124;
Matches  267;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy  56 TCCGGGTTCGGCGCGGCGGCGGCGGGGATGGAGTGAATCCCGATGGAGCGCCCGAGGAGGCAAGC 115
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5 TCCGGGTTCGGCGCGGCGGCGGCGGGGATGGAGTGAATCCCGATGGAGCGCCCGAGGAGGCAAGC 64

Qy 116 AGTCGCCGCGCGCCAGCCCTGGGGACGGCTCCTCGCTCTGCTCGCTCTGCGTCCTGGGCGCGGAGGGCGAGGGGCGAGC 175
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db  65 AGTCGCCGCGCGCCAGCCCTGGGGACGGCTCCTCGCTCTGCTCGCTCTGCGTCCTGGGCGCGAGAGCGAGGGGCGAGC 124

Qy 176 CGCACGTCCTCCTGAGGAAGCGGGAGTGGACCATCGGGCGCGGAGACGAGGTTGCGACCTTT 235
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 125 CGCACGTCCTCCTGAGGAAGCGGGAGTGGACCATCGGGCGCGGAGACGAGGTTGCGACCTTT 184

Qy 236 CCTTCCCAGCAATAAACTGGTCTCTGAGAGATCACTGTAGAATTGTAGTGGATGAAAAAT 295
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 185 CCTTCCCAGCAATAAACTGGTCTCTGAGAGATCACTGTAGAATTGTAGTGGATGAAAAAT 244

Qy 296 CAGGTCAGGTGACACTGGAAGATACCA 322
        |||||||||||||||||||||||||||
Db 245 CAGGTCAGGTGACACTGGAAGATACCA 271


RESULT 7
US-09-764-864-323
; Sequence 323, Application US/09764864
; Patent No. US2002013753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 323
; LENGTH: 354
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (289)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (313)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-323

Query Match     8.9%;  Score 238;  DB 9;  Length 354;
Best Local Similarity  100.0%;  Pred. No. 4.3e-109;
Matches  238;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy 245 GCAATAAACTGGTCTCTGAGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGTCAGG 304
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  51 GCAATAAACTGGTCTCTGAGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGTCAGG 110

Qy 305 TGACACTGGAAGATACCAGCACCAGTGGAACAGTGATTAACAGCTGAGGTTGTTAAGA 364
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111 TGACACTGGAAGATACCAGCACCAGTGGAACAGTGATTAACAGCTGAGGTTGTTAAGA 170

Qy 365 AGCAGACATGCCCTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAATGAAC 424
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 171 AGCAGACATGCCCTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAATGAAC 230

Qy 425 CGGAACACACGTGCATACCTCATGAATCTTTAAGTGAAAAGCAAGGCATGACACA 482
        |||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db 231 CGGAACACACGTGGCATACCTCATGAATCTTTAAGTGAAAAGCAAGGCATGACACA 288


RESULT 8
US-09-764-864-742
; Sequence 742, Application US/09764864
; Patent No. US2002013753A1
; GENERAL INFORMATION:
```

```
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PT223
;   CURRENT APPLICATION NUMBER: US/09/764,864
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 1792
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 742
;   LENGTH: 354
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (289)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (313)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-742

  Query Match             8.9%;  Score 238;  DB 9;  Length 354;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-109;
  Matches  238;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    245 GCAATAAACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGTCAGG 304
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51 GCAATAAACTGGTCTCTGGAGATCACTGTAGAATTGTAGTGGATGAAAAATCAGGTCAGG 110

Qy    305 TGACACTGGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGA 364
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    111 TGACACTGGAAGATACCAGCACCAGTGGAACAGTGATTAACAAGCTGAAGGTTGTTAAGA 170

Qy    365 AGCAGACATGCCCTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAATGAAC 424
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    171 AGCAGACATGCCCTTTACAGACTGGGGATGTCATCTACTTGGTGTACAGGAAGAATGAAC 230

Qy    425 CGGAACACAACGTGGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGGCATGACACA 482
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 CGGAACACAACGTGGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGGCATGACACA 288

RESULT 9
US-09-918-995-13662
;   Sequence 13662, Application US/09918995
;   Publication No. US20030073623A1
;   GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;   TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;   FILE REFERENCE: 20411-756
;   CURRENT APPLICATION NUMBER: US/09/918,995
;   CURRENT FILING DATE: 2001-07-30
;   PRIOR APPLICATION NUMBER: US/09/235,076
;   PRIOR FILING DATE: 1999-01-20
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   NUMBER OF SEQ ID NOS: 38054
;   SEQ ID NO 13662
;   LENGTH: 449
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(449)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-13662

  Query Match             8.1%;  Score 217;  DB 10;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-98;
  Matches  217;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   2445 AGTTTCAAGGGAGGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAG 2504
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     92 AGTTTCAAGGGAGGAAGGGCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAG 151

Qy   2505 TTTACAGCCTACAGACGTACCACAATATCCTGCTGCTGGGAAAAACCACAGCATTTTATCT 2564
          |||||||||||||| ||||| |||||||||||||| |||||||||||||||||||||||||
Db    152 TTTACAGCCTACAGGACGTACAACAATATCCTGCTGGGGAAAACCACAGCATTTTATCT 211

Qy   2565 ATTTTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAGAATTTAAATGTCACAAACTG 2624
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    212 ATTTTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAGAATTTAAATGTCACAAACTG 271

Qy   2625 TAGCACAAATATATAATATTATAATTTACAAATTGAC 2661
          |||||||||||||||||||||||||||||||||||||
Db    272 TAGCACAAATATATAATATTATAATTTACAAATTGAC 308

RESULT 10
US-09-918-995-2180
;   Sequence 2180, Application US/09918995
;   Publication No. US20030073623A1
;   GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;   TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;   FILE REFERENCE: 20411-756
;   CURRENT APPLICATION NUMBER: US/09/918,995
;   CURRENT FILING DATE: 2001-07-30
;   PRIOR APPLICATION NUMBER: US/09/235,076
;   PRIOR FILING DATE: 1999-01-20
;   NUMBER OF SEQ ID NOS: 38054
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 2180
;   LENGTH: 476
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(476)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-2180

  Query Match             2.9%;  Score 77;  DB 10;  Length 476;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-28;  Indels   0;  Gaps   0;
  Matches   77;  Conservative   0;  Mismatches   0;

Qy   2046 ATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCA 2105
          |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     52 ATTCAATCATATCTGTGAACAGACAAGGTTCAAAAACTAAGCATCCAGAGAGCCCTGAGCA 111

Qy   2106 GCTTTCAGCACTGGAGG 2122
          |||||||||||||||||
Db    112 GCTTTCAGCACTGGAGG 128

RESULT 11
US-09-908-975-6220
;   Sequence 6220, Application US/09908975
;   Publication No. US20030165843A1
;   GENERAL INFORMATION:
;   APPLICANT: SHOSHAN, Avi
;   APPLICANT: WASSERMAN, Alon
;   APPLICANT: MINTZ, Eli
;   APPLICANT: MINTZ, Liat
;   APPLICANT: FAIGLER, Simchon
;   TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE \
;   TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
;   FILE REFERENCE: 36688-0005
;   CURRENT APPLICATION NUMBER: US/09/908,975
;   CURRENT FILING DATE: 2001-07-20
;   PRIOR APPLICATION NUMBER: US 60/287,724
;   PRIOR FILING DATE: 2001-05-02
;   PRIOR APPLICATION NUMBER: US 60/221,607
;   PRIOR FILING DATE: 2000-07-28
;   NUMBER OF SEQ ID NOS: 32337
```

```
;       SOFTWARE: PatentIn version 3.0
;  SEQ ID NO 6220
;       LENGTH: 60
;       TYPE: DNA
;       ORGANISM: Homo sapiens
US-09-908-975-6220

Query Match             2.2%;  Score 60;  DB 10;  Length 60;
Best Local Similarity   100.0%;  Pred. No. 2.8e-19;
Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1843 GGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGT 1902
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACACCGTTCTGTGTTACTGCTGT 60


RESULT 12
US-10-304-928-15
;  Sequence 15, Application US/10304928
;  Publication No. US20030101484A1
;  GENERAL INFORMATION:
;       APPLICANT: Simmons, Carl R.
;       APPLICANT: Yalpani, Nasser
;       TITLE OF INVENTION: Maize Chitinases and Their Use in
;       TITLE OF INVENTION: Enhancing Disease Resistance in Crop Plants
;       FILE REFERENCE: 1100D
;       CURRENT APPLICATION NUMBER: US/10/304,928
;       CURRENT FILING DATE:  2002-11-26
;       PRIOR APPLICATION NUMBER: 09/522,714
;       PRIOR FILING DATE: 2000-03-10
;       PRIOR APPLICATION NUMBER: 60/125,915
;       PRIOR FILING DATE: 1999-03-24
;       NUMBER OF SEQ ID NOS: 32
;       SOFTWARE: FastSEQ for Windows Version 3.0
;  SEQ ID NO 15
;       LENGTH: 1085
;       TYPE: DNA
;       ORGANISM: Zea mays
;       FEATURE:
;       NAME/KEY: CDS
;       LOCATION: (46)...(966)
US-10-304-928-15

Query Match             0.9%;  Score 25;  DB 15;  Length 1085;
Best Local Similarity   100.0%;  Pred. No. 0.13;
Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2655 AATTGACAAAAAAAAAAAAAAAAAAAA 2679
              |||||||||||||||||||||||||
Db       1061 AATTGACAAAAAAAAAAAAAAAAAAAA 1085


RESULT 13
US-09-957-664-1
;  Sequence 1, Application US/09957664
;  Patent No. US2002012309741
;  GENERAL INFORMATION:
;       APPLICANT: Curtis, Rory A. J.
;       TITLE OF INVENTION: 63760, A NOVEL HUMAN TRANSPORTER AND
;       TITLE OF INVENTION: USES THEREFOR
;       FILE REFERENCE: MNI-192
;       CURRENT APPLICATION NUMBER: US/09/957,664
;       CURRENT FILING DATE: 2001-09-19
;       PRIOR APPLICATION NUMBER: 60/233,790
;       PRIOR FILING DATE: 2000-09-19
;       NUMBER OF SEQ ID NOS: 4
;       SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 1
;       LENGTH: 1963
;       TYPE: DNA
;       ORGANISM: Homo sapiens
;       FEATURE:
;       NAME/KEY: CDS
```

```
;       LOCATION: (67)...(1491)
US-09-957-664-1

Query Match             0.9%;  Score 25;  DB 9;  Length 1963;
Best Local Similarity   100.0%;  Pred. No. 0.13;
Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2655 AATTGACAAAAAAAAAAAAAAAAAAAA 2679
              |||||||||||||||||||||||||
Db       1916 AATTGACAAAAAAAAAAAAAAAAAAAA 1940


RESULT 14
US-10-154-419-15
;  Sequence 15, Application US/10154419
;  Publication No. US20030143675A1
;  GENERAL INFORMATION:
;       APPLICANT: Curtis, Rory A.J.
;       APPLICANT: Glucksman, Maria Alexandra
;       APPLICANT: Meyers, Rachel E.
;       TITLE OF INVENTION: NOVEL 38594, 57312, 53659, 57250, 63760, 49938, 32146,
;       TITLE OF INVENTION: 57259, 57118, 67067, 62092, 8099, 46455, 54414, 53763,
;       TITLE OF INVENTION: 67076, 67102, 44181, 67084FL, 67084ALT, FBH58295FL, 57255,
;       TITLE OF INVENTION:  AND 57255alt MOLECULES AND USES THEREFOR
;       FILE REFERENCE: MNI-249
;       CURRENT APPLICATION NUMBER: US/10/154,419
;       CURRENT FILING DATE:  2002-05-22
;       Prior Application removed - See Palm or File Wrapper
;       NUMBER OF SEQ ID NOS: 99
;       SOFTWARE: FastSEQ Version 4.0
;  SEQ ID NO 15
;       LENGTH: 1963
;       TYPE: DNA
;       ORGANISM: Homo sapiens
;       FEATURE:
;       NAME/KEY: CDS
;       LOCATION: (67)...(1491)
US-10-154-419-15

Query Match             0.9%;  Score 25;  DB 15;  Length 1963;
Best Local Similarity   100.0%;  Pred. No. 0.13;
Matches   25;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2655 AATTGACAAAAAAAAAAAAAAAAAAAA 2679
              |||||||||||||||||||||||||
Db       1916 AATTGACAAAAAAAAAAAAAAAAAAAA 1940


RESULT 15
US-09-938-842A-3337/c
;  Sequence 3337, Application US/09938842A
;  Patent No. US2002016037841
;  GENERAL INFORMATION:
;       APPLICANT: Harper, Jeff
;       APPLICANT: Kreps, Joel
;       APPLICANT: Wang, Xun
;       APPLICANT: Zhu, Tong
;       TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;       TITLE OF INVENTION: SAME, AND METHODS OF USE
;       FILE REFERENCE: SCRIP1300-3
;       CURRENT APPLICATION NUMBER: US/09/938,842A
;       CURRENT FILING DATE:  2001-08-24
;       PRIOR APPLICATION NUMBER: US 60/227,866
;       PRIOR FILING DATE: 2000-08-24
;       PRIOR APPLICATION NUMBER: US 60/264,647
;       PRIOR FILING DATE: 2001-01-16
;       PRIOR APPLICATION NUMBER: US 60/300,111
;       PRIOR FILING DATE: 2001-06-22
;       NUMBER OF SEQ ID NOS: 5379
;  SEQ ID NO 3337
;       LENGTH: 2000
;       TYPE: DNA
;       ORGANISM: Arabidopsis thaliana
```

US-09-938-842A-3337

```
Query Match           0.9%;  Score 25;  DB 9;  Length 2000;
Best Local Similarity 100.0%;  Pred. No. 0.13;
Matches   25;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     2655 AATTGACAAAAAAAAAAAAAAAAA 2679
            |||||||||||||||||||||||||
Db     1268 AATTGACAAAAAAAAAAAAAAAAA 1244

Search completed: May 15, 2004, 12:03:05
Job time : 1595 secs
```

```
                    Copyright   GenCore version 5.1.6
                    Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:       May 14, 2004, 23:36:46 ; Search time 10039 seconds
                                        (without alignments)
                                        7968.999 Million cell updates/sec

Title:          US-10-048-046-1
Perfect score:  2679
Sequence:       1 aagaattcggcacgaggccg.........acaaaaaaaaaaaaaaaaaa 2679

Scoring table:  OLIGO_NUC
                Gapop 60.0 , Gapext 60.0

Searched:       27513289 seqs, 14931090276 residues

Word size :     0

Total number of hits satisfying chosen parameters:      55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :      EST:*
                1:  em_estba:*
                2:  em_esthum:*
                3:  em_estin:*
                4:  em_estmu:*
                5:  em_estov:*
                6:  em_estpl:*
                7:  em_estro:*
                8:  em_htc:*
                9:  gb_est2:*
               10:  gb_est3:*
               11:  gb_htc:*
               12:  gb_est3:*
               13:  gb_est4:*
               14:  gb_est5:*
               15:  em_estfun:*
               16:  em_estom:*
               17:  em_gss_hum:*
               18:  em_gss_inv:*
               19:  em_gss_pln:*
               20:  em_gss_vrt:*
               21:  em_gss_fun:*
               22:  em_gss_mam:*
               23:  em_gss_mus:*
               24:  em_gss_pro:*
               25:  em_gss_rod:*
               26:  em_gss_phg:*
               27:  em_gss_vrl:*
               28:  gb_gss1:*
               29:  gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	825	30.8	1201	9	AL539798	AL539798 AL539798
2	722	27.0	921	13	BU176559	BU176559 AGENCOURT
3	703	26.2	992	13	BQ073354	BQ073354 AGENCOURT
4	698	26.1	877	14	CD243458	CD243458 AGENCOURT

No.	Score	%	Len	N	ID1	ID2
5	647	24.2	897	10	BF971929	BF971929 602240332
6	620	23.1	840	12	BG395714	BG395714 602453365
7	600	22.4	680	9	AU134937	AU134937 AU134937
8	583	21.8	870	13	BQ215731	BQ215731 AGENCOURT
9	565	21.1	916	12	BM459647	BM459647 AGENCOURT
10	563	21.0	1143	12	BM542545	BM542545 AGENCOURT
11	560	20.9	666	10	BE407732	BE407732 60129777
12	539	20.1	1201	9	AL539797	AL539797 AL539797
13	519	19.4	934	10	BF971838	BF971838 602240424
14	513	19.1	729	14	CA916737	CA916737 30LSPRIM2
15	488	18.2	690	10	BE280165	BE280165 60115370
16	478	17.8	910	13	BU148471	BU148471 AGENCOURT
17	478	17.8	938	13	BQ887242	BQ887242 AGENCOURT
18	477	17.8	1092	12	BM541730	BM541730 AGENCOURT
19	469	17.5	909	13	BX409791	BX409791 BX409791
20	468	17.5	824	9	AU132825	AU132825 AU132825
21	456	17.0	712	10	BE257482	BE257482 601111439
22	456	17.0	1096	12	BG335362	BG335362 602404192
23	452	16.9	1022	12	BM558904	BM558904 AGENCOURT
24	450	16.8	953	13	BQ936072	BQ936072 AGENCOURT
25	447	16.7	938	13	BQ924736	BQ924736 AGENCOURT
26	436	16.3	535	13	BX486289	BX486289 DKFZp686P
27	436	16.3	557	13	BX486564	BX486564 DKFZp686H
28	435	16.2	811	13	CB989147	CB989147 AGENCOURT
29	433	16.2	806	14	CB961597	CB961597 AGENCOURT
30	426	15.9	518	9	AU154212	AU154212 AU154212
31	421	15.7	687	9	AU138310	AU138310 AU138310
32	420	15.7	876	13	BQ876252	BQ876252 AGENCOURT
33	418	15.6	650	10	AW957709	AW957709 EST369779
34	412	15.4	635	12	BG481936	BG481936 602526853
35	409	15.3	490	10	BF109790	BF109790 7169a04.x
36	404	15.1	881	13	BU153737	BU153737 AGENCOURT
37	394	14.7	944	13	BX409790	BX409790 BX409790
38	386	14.4	475	9	AI200393	AI200393 qf86e04.x
39	382	14.3	824	12	BG831289	BG831289 602766120
40	381	14.2	797	12	BI460002	BI460002 603201532
41	364	13.6	1070	10	BM803324	BM803324 AGENCOURT
42	354	13.2	636	10	BE407797	BE407797 601299183
43	335	12.5	697	12	BG473597	BG473597 602515946
44	326	12.2	527	12	BF439206	BF439206 nab60h05.
45	320	11.9	722	9	AU135780	AU135780 AU135780
```

ALIGNMENTS

```
RESULT 1
AL539798
LOCUS       AL539798            1201 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL539798 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF031YA23 5-PRIME, mRNA sequence.
ACCESSION   AL539798
VERSION     AL539798.2  GI:31264361
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12869349.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr; Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6792.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF031AA12QP1&cluster=6792.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
```

```
http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DF031AA12QP1.
FEATURES
     source          1..1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF031YA23"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primeI. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

Query Match     30.8%;  Score 825;  DB 9;  Length 1201;
Best Local Similarity   100.0%;  Pred. No. 1.4e-155;
Matches  825;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps    0;

Qy     61 GTTCGGGGCGGGGCGGGGGATGTGAATCCGATGGAGCGGCCCGAGGAGGCAAGCAGTCG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     67 GTTCGGGGCGGGGCGGGGGATGTGAATCCGATGGAGCGGCCCGAGGAGGCAAGCAGTCG 126

Qy    121 CCGCCGCCCAGCCCTGGGGACGGCTCCTCGCGTCTGGGCGGCGGAGGGCGAGCCGCAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    127 CCGCCGCCGAGCCCTGGGGACGGCTCCTCGCGTCTGGGCGGCGGAGGGCGAGCCGCAC 186

Qy    181 GTCCTCCTGAGGAAGCGGGAGTGGACCATCGGGCGGCGGAGTGAGACGAGGTTGCGACCTTTCCTTC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 GTCCTCCTGAGGAAGCGGGAGTGGACCATCGGGCGGCGGAGTGAGACGAGGTTGCGACCTTTCCTTC 246

Qy    241 CCCAGCAATAAACTGGTCTCTGAGATCACTGTAGAATTGTAGTGAGAAAAATCAGGT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247 CCCAGCAATAAACTGGTCTCTGAGATCACTGTAGAATTGTAGTGAGAAAAATCAGGT 306

Qy    301 CAGGTGACACTGAAGATACCAGCACCAGTGAACACAGTGATTAACAGTCTGAAGGTTGTT 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 CAGGTGACACTGAAGATACCAGCACCAGTGAACACAGTGATTAACAGTCTGAAGGTTGTT 366

Qy    361 AAGAAGCAGACACATGCCCTTTACAGACTGGGGATGTCATCTCACTGGTGTACAGGAAGAAT 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    367 AAGAAGCAGACACATGCCCTTTACAGACTGGGGATGTCATCTCACTGGTGTACAGGAAGAAT 426

Qy    421 GAACCGGAACACAACGTGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGGCATGACA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    427 GAACCGGAACACAACGTGCATACCTCTATGAATCTTTAAGTGAAAAGCAAGGCATGACA 486

Qy    481 CAAGAATCCTTTGAAGCTAACAAGGAAAAATGTGTTCCATGGGACCAAAGATACCTCAGGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    487 CAAGAATCCTTTGAAGCTAACAAGGAAAAATGTGTTCCATGGGACCAAAGATACCTCAGGT 546

Qy    541 GCAGGTGCAGGGGCAGGGGCCGATCCCCGGTCCCCGGTCCCCGGGCCCACTCAGGTG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    547 GCAGGTGCAGGGGCAGGGGCCGATCCCCGGTCCCCGGTCCCCGGGCCCACTCAGGTG 606

Qy    601 TGCTTTGAGGAAGCACCACAGCCATCAACATCGACGTCAGAACCTCTTCCCCACAGCCTCGGCC 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    607 TGCTTTGAGGAAGCACCACAGCCATCAACATCGACGTCAGAACCTCTTCCCCACAGCCTCGGCC 666

Qy    661 TCTTCCACGGAGGCCTTCTCCTCGAGGGGCGGAGAGCGGTTCCTCCAGTGTGGGTGCTCGGGGGT 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    667 TCTTCCACGGAGGCCTTCTCCTCGAGGGGCGGAGAGCGGTTCCTCCAGTGTGGGTGCTCGGGGGT 726

Qy    721 GGTGGCATCTCCCCTAAAGGAAGTGGTCCTCTGTGGCAAGTGATGAAGTCTTCCAGCTTT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    727 GGTGGCATCTCCCCTAAAGGAAGTGGTCCTCTGTGGCAAGTGATGAAGTCTTCCAGCTTT 786

Qy    781 GCCTCAGCTTCTCCCAGACAGAAAGAGACTGCGTCCTTTTCGTCGTTGGAACCCCAGGATCAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    787 GCCTCAGCTTCTCCCAGACAGAAAGAGACTGCGTCCTTTTCGTCGTTGGAACCCCAGGATCAG 846
```

```
Qy    841 GAGGATTTGGAGCCCGTGAAGAAGAAAATGAGAGGAGGATGGGGAC 885
          |||||||||||||||||||||||||||||||||||||||||||||
Db    847 GAGGATTTGGAGCCCGTGAAGAAGAAAATGAGAGGAGGATGGGGAC 891


RESULT 2
BU176559
LOCUS       BU176559                 921 bp    mRNA      linear  EST 04-SEP-2002
DEFINITION  AGENCOURT_7940317 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6143609
            5', mRNA sequence.
ACCESSION   BU176559
VERSION     BU176559.1  GI:22690543
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13467 row: e  column: 18
            High quality sequence stop: 697.
FEATURES
     source          1..921
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6143609"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_67"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally. Primer: oligo dT.
                     Average insert size 1.75 kb.  Library constructed by Life
                     Technologies."
ORIGIN

Query Match     27.0%;  Score 722;  DB 13;  Length 921;
Best Local Similarity   99.9%;  Pred. No. 5.5e-135;
Matches  772;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps    0;

Qy   1889 TGTGTTACTGCGTGTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACA 1948
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGTGTTACTGCGTGTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACA 60

Qy   1949 TTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCCGTCCGTCTGGGGCCGTA 2008
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCCGTCCGTCTGGGGCCGTA 120

Qy   2009 ACTGCGGCACTCAGGTGAAAGCTCACCAGCCATCAGAAATTCAATCATATCTGTGAACAGA 2068
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACTGCGGCACTCAGGTGAAAGCTCACCAGCCATGAAATTCAATCATATCTGTGAACAGA 180

Qy   2069 CAAGGTTCAAAAACTAAGCATCCAAGGCCCTGAGCAGCTTTCAGCACTGAAGGTGAAGA 2128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAGGTTCAAAAACTAAGCATCCAAGAGCCCTGAGCAGCTTTCAGCACTGAAGGTGAAGA 240

Qy   2129 GAGCGTGTTTTAAAATACAGAGACAAGCAGCACGTCAAGGTGTTTTCACAGCCCCTGAGGG 2188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGCGTGTTTTAAAATACAGAGACAAGCAGCACGTCAAGGTGTTTTCACAGCCCCTGAGGG 300

Qy   2189 AAGGGACGCAGGGTCTCCGACAGGTGTCTCGGGGTGACTCTTCTGTGGGAGCTTTTTACCC 2248
```

Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN

Query Match   26.2%; Score 703; DB 13; Length 992;
Best Local Similarity 100.0%; Pred. No. 3.1e-131;
Matches 703; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    551 GGGGAGGGGCGGATCCCCGGGTCCTCGTCGCCGGCCACTCAGGTGTGCTTTGAGG  610
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGGGAGGGGCGGATCCCCGGGTCCTCGTCGCCGGCCACTCAGGTGTGCTTTGAGG   60

Qy    611 AACCACAGCCATCAACATCGACGTCAGACCTCTTCCCACAGCCTCGGCCTCTTCCACGG  670
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AACCACAGCCATCAACATCGACGTCAGACCTCTTCCCACAGCCTCGGCCTCTTCCACGG  120

Qy    671 AGCCTTCTCCTGCAGGGGCGAGAGCGTTCCTCCCAGTGTGGGTCTGGGGTCTGGGTGGTGGCATCT  730
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGCCTTCTCCTGCAGGGGCGAGAGCGTTCCTCCCAGTGTGGGTCTGGGGTCTGGGTGGTGGCATCT  180

Qy    731 CCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTTGCCTCAGCTC  790
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCCCTAAAGGAAGTGGTCCCTCTGTGGCAAGTGATGAAGTCTCCAGCTTTGCCTCAGCTC  240

Qy    791 TCCCAGACGAAAGACTGCGTCCTTTTGCGTCCTTTTCGTGGTCGTGGAACCCCAGGATCAGGAGGATTTGG  850
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCCCAGACGAAAGACTGCGTCCTTTTGCGTCCTTTTCGTGGTCGTGGAACCCCAGGATCAGGAGGATTTGG  300

Qy    851 AGCCCGTGAAGAAGAAAATGAGAGGAGATGGGGAGATGGGGACCTTGACCTGAACCGGCAGTGTTGG  910
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCCCGTGAAGAAGAAAATGAGAGGAGATGGGGAGATGGGGACCTTGACCTGAACCGGCAGTGTTGG  360

Qy    911 TCGCACAACCGGTAGAAATGCCCAAACCGTCCAGGAGACGTCAGAGGACGTCAGACGCGGCGGCTGGGA  970
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCGCACAACCGGTAGAAATGCCCAAACCGTCCAGGAGACGTCAGAGGACGTCAGACGCGGCGGCTGGGA  420

Qy    971 AGCCAGAACAAGATGCGAGGAGACGCTGACATGCACACGTTCTGCGCGGCTTGCTACTCGGGCGTGGATG  1030
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGCCAGAACAAGATGCGAGGAGACGCTGACATGCACACGTTCTGCGCGGCTTGCTACTCGGGCGTGGATG  480

Qy   1031 GCGGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCCTTGCTACTCGGGCGTGGATGG  1090
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCGGTGAGTTTGCAGCCCTGCATGCACACGTTCTGCGCCTTGCTACTCGGGCGTGGATGG  540

Qy   1091 AGGGCTCGTCCCGTGTCCTACTGCCGTGGAGCCGGATCTGTAAAAACCACA  1150
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGCGGCTCGTCCCGTGTCCTACTGCCGTGGAGCCGGATCTGTAAAAACCACA  600

Qy   1151 TCCTCAACAACCTCGTGGAAGCATACCTCATCCAGACAAGAGTCAGCAGTCGCAGTGAAG  1210
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TCCTCAACAACCTCGTGGAAGCATACCTCATCCAGACAAGAGTCAGCAGTCGCAGTGAAG  660

Qy   1211 AAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGA  1253
          |||||||||||||||||||||||||||||||||||||||||
Db    661 AAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGA  703
```

RESULT 4
CD243458
CD243458
LOCUS   CD243458   877 bp   mRNA   linear   EST 22-MAY-2003
DEFINITION AGENCOURT 14121909 NIH_MGC 180 Homo sapiens cDNA clone
            IMAGE:30383176 5', mRNA sequence.
ACCESSION   CD243458
VERSION   CD243458.1 GI:31003922
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 877)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)

```
Db    301 AAGGGACGCAGGGTCCCCGGACCAGGGTCGCTCTGGGGGTGACTCTTCTGTGGAGCTTTTACCC  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   2249 TCTGAGTGAGACCCTCCCCAGAGCCCGGGGGCCCGCACCCCTCCTGGTGAGCGCGTG  2308
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCTGAGTGAGACCCTCCCCAGAGCCCGGGGGCCCGCACCCCTCCTGGTGAGCGCGTG  420

Qy   2309 GGCAGGGCTCGTGGTGGCATCAGCAGCAGAGACAGAGCAGAAGCCTTCTCTGTAACATGCGGCCGTC  2368
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGCAGGGCTCATGGTGGCATCAGCAGCAGAGACAGAGCAGAAGCCTTCTCTGTAACATGCGGCCGTC  480

Qy   2369 CCGCCAGAGAGGGCAGTTTTGCTCTTTTGTACATTTCCGAAACTACACAGTTAAAGCAGAA  2428
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCGCCCAGAGAGGGCAGTTTTGCTCTTTTGTACATTTCCGAAACTACACAGTTAAAGCAGAA  540

Qy   2429 GTCGTGTTTCAGGAAAAGTTCAAGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAG  2488
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GTCGTGTTTCAGGAAAAGTTCAAGGAGAAGGGCAAGTTTATCAAAAACATTGTTTCAG  600

Qy   2489 GAGAAGGGGAGCCATAAGTTTACAGCCTACAGGAGCGTACACAATATCCTGCTGGGAAAA  2548
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAGAAGGGGAGCCATAAGTTTACAGCCTACAGGAGCGTACACAATATCCTGCTGGGAAAA  660

Qy   2549 CCACAGCCATTTATCATCTATTTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGAATT  2608
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CCACAGCCATTTATCTATTTTTTTATTTTAATAGGTTTGGTGCTTATCTTCTAATAAGAATT  720

Qy   2609 TAAATGTCACAAACTGTAGCACAAATAATAATTAATTTATAATTTACAAATTGAC  2661
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TAAATGTCACAAACTGTAGCACAAATAATAATTAATTTATAATTTACAAATTGAC  773
```

RESULT 3
BQ073354
BQ073354
LOCUS   BQ073354   992 bp   mRNA   linear   EST 02-APR-2002
DEFINITION AGENCOURT_7052199 NIH_MGC_101 Homo sapiens cDNA clone IMAGE:5806143
            5', mRNA sequence.
ACCESSION   BQ073354
VERSION   BQ073354.1 GI:19902400
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 992)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2046 row: h column: 16
            High quality sequence stop: 670.
FEATURES   Location/Qualifiers
     source   1..992
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5806143"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_101"
                     /note="Organ: lung; Vector: pOTB7; Site_1: EcoRI; Site_2:
                     XhoI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,

FEATURES
     source   1..877
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30383176"
                     /lab_host="DH10B-TonA (T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC_180"
                     /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site_1: NotI;
                     Site_2: EcoRV (destroyed); Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.68 kb. Library was
                     constructed by (Invitrogen). Note: this is a NIH_MGC
                     Library."
ORIGIN

Query Match   26.1%;   Score 638; DB 14; Length 877;
Best Local Similarity 99.9%; Pred. No. 3.4e-130;
Matches 748; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

Qy 1527 AGGAAGCCACGCCCTGCACCTGCTTCGTGCTTCCAGCCCATGCCCGACCCGGAGAGGCGGAGCG 1586
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4 AGGAAGCCACGCCCTGCACCTGCTTGCTTCCAGCCCATGCCCGACCCGGAGAGGCGGAGCG 63

Qy 1587 CGAGCAGGAGGACCCGGCGTGTCGCCCCTCAGCAGTGTGCGGTCTCGCTGCAGCCTTTCTGCCA 1646
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   64 CGAGCAGGAGGACCCGGCGTGTCGCCCCTCAGCAGTGTGCGGTCTCGCTGCAGCCTTTCTGCCA 123

Qy 1647 CCTGTACTGGGGGCTGCACCCGGACCGGCTGCTACGGCTGCCTGGCCCCGTTTTGTGAGCT 1706
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   124 CCTGTACTGGGGGCTGCACCCGGACCGGCTGCTACGGCTGCCTGGCCCCGTTTTGTGAGCT 183

Qy 1707 CAACCTGGGTGACAAGTGTCTGGACGGCGGTGCTGAACAACAACAGCTACGAGTGCAGAGAT 1766
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 CAACCTGGGTGACAAGTGTCTGGACGGCGGTGCTGAACAACAACAGCTACGAGTGCAGAGAT 243

Qy 1767 CCTGAAGAATTACCTGGACAACCAGAGGTTTGACATGAACATGTTGACCGAGAGCT 1826
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244 CCTGAAGAATTACCTGGCAACCAGAGGTTTGACATGAACATGTTGACCGAGAGCCT 303

Qy 1827 CGTGGCTCTCCAGCGGGGAGTGTTTTCTGCTGTCTGATTACAGAGTCACGGAGACACCGT 1886
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   304 CGTGGCTCTCCAGCGGGGACGTGTTTTCTGCTGTCTGATTACAGAGTCACGGAGACACCGT 363

Qy 1887 TCTGTGTTACTGCTGTGGCCTGCGCAGCTTCCGTGAGCGAGCTGACCTAGCAGTATCGGCAGAA 1946
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   364 TCTGTGTTACTGCTGTGGCCTGCGCAGCTTCCGTGAGCGAGCTGACCTAGCAGTATCGGCAGAA 423

Qy 1947 CATTCCTGCTTCCGAGTTGCCAGTGGCCGTAAACATCCCGTCGACTGCTACTGGGGCCG 2006
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   424 CATTCCTGCTTCCGAGTTGCCAGTGGCCGTAAACATCCCGTCCTGACTGCTACTGGGGCCG 483

Qy 2007 TAACTGCCGCACTCAGGTGAAAGTCACCACGCCATGAAATTCAATCATCATATCGTGAACA 2066
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   484 TAACTGCCGCACTCAGGTGAAAGTCACCACGCCATGAAATTCAATCATCATATCGTGAACA 543

Qy 2067 GACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAGGTGAA 2126
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   544 GACAAGGTTCAAAAACTAAGCATCCAGAGGCCCTGAGCAGCTTTCAGCACTGGAGGTGAA 603

Qy 2127 GAGAGGGTGTGTTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTTCACAGCCCCCTGAG 2186
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   604 GAGAGGGTGTGTTTTTAAAATACAGAGACAAGCACGTCAAGGTGTTTTCACAGCCCCCTGAG 663
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2187 GGAAGGGACGCAGGGTCTCCGACAGTGCTCTCGGGGTGACTCTTCTGTGTGGAGCTTTTTAC 2246
         ||||||||||||||||||||||||||
Db   664 GGAAGGGACGCAGGNTCTCCCACAGGGTGCTCTGGGGTGACTCTTCTGTGTGGAGCTTTTAC 723

Qy 2247 CCTCTGAGTGAGACCCTCCCCAGAGCCC 2275
         ||||||||||||||||||||||||
Db   724 CCTCTGAGTGAGACCCTCCCCAGAGCCCC 752

RESULT 5
FEATURES   Location/Qualifiers
     source   1..897
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4329079"
                     /tissue_type="leiomyosarcoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_46"
                     /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

Query Match   24.2%;   Score 647; DB 10; Length 897;
Best Local Similarity 100.0%; Pred. No. 4.7e-120;
Matches 647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1908 GCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCC 1967
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCC 60

Qy 1968 AGTGGCCGTAACATCCCGTCCTACTGCTACTGCGGGGCCGTAACTGCCGCACTCAGGTGAA 2027
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AGTGGCCGTAACATCCCGTCCTACTGCTACTGCGGGGCCGTAACTGCCGCACTCAGGTGAA 120

Qy 2028 AGTCACCACGCCATGAAATTCAATCATCATATCGTGAACAGACAAGGTTCAAAAACTAAGC 2087
         |||||||||||||||||||||||||||||||||||||||||||||||||||||

adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

```
Query Match        23.1%;   Score 620;  DB 12;   Length 840;
Best Local Similarity 99.9%;  Pred. No. 1.2e-114;
Matches  670;  Conservative 0;  Mismatches 1;  Indels 1;  Gaps 0;

Qy  1294 GAAGGGAGTTCAGAGGACCTGCTGGAGCTGTCAGACGTTGACAGTGAGTCCTCAGACATT 1353
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2 GAAGGGAGTTCAGAGGACCTGCTGGAGCTGTCAGACGTTGACAGTGAGTCCTCAGACATT   61

Qy  1354 AGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCGGCCGCAGCCT 1413
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62 AGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGAAGGCAGGCGGCCGCAGCCT  121

Qy  1414 CCCCACTGCCCAGCACCCGAGGGCGAGCCCACAGGCCCTGGGGGATGCACCC 1473
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 CCCCACTGCCCAGCACCCGAGGGCGAGCCCACAGGCCCTGGGGGATGCACCC  181

Qy  1474 TCCACGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTGTGCCCCTGCAAGGAAGC 1533
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 TCCACGTCCGTCAGCCTGACGACAGCAGTCCAGGATTACGTGTGCCCCTGCAAGGAAGC  241

Qy  1534 CACGCCCCTGTGCACCTGCCTTCCAGCCCATGCCCGACCGGAGACGGGAGGCGGAGCAG 1593
          |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db   242 CACGCCCCTGTGCACCTGCCTTCCAGCCCATGCCCGACCGGAGAGTGGAGGCGGAGCAG  301

Qy  1594 GACCCGCGTGTCGCCCCTCAGCAGTGTGCGGTCCTGCCTCGCAGCCTTTCTGCCACCCTAC 1653
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 GACCCGCGTGTCGCCCCTCAGCAGTGTGCGGTCCTGCCTCGCAGCCTTTCTGCCACCCTAC  361

Qy  1654 TGGGGCTGCACCCGGACCGGCTGCTACGGCTGCTGCTGTCCGGCCCGTTTGTGAGCTCAACCTG 1713
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362 TGGGGCTGCACCCGGACCGGCTGCTACGGCTGCTGCTGTCCGGCCCGTTTGTGAGCTCAACCTG  421

Qy  1714 GGTGACAAGTGTCTGGACGGCGCTGCTGAACAACAACAGCTACGAGTCAGAGATCCTGAAG 1773
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422 GGTGACAAGTGTCTGGACGGCGCTGCTGAACAACAACAGCTACGAGTCAGAGATCCTGAAG  481

Qy  1774 AATTACCTGGCAACCAGAGGTTGACATGGGAAAAACATGTTGACCGGAGAGCCTCGTGCT 1833
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   482 AATTACCTGGCAACCAGAGGTTGACATGGGAAAAACATGTTGACCGGAGAGCCTCGTGCT  541

Qy  1834 CTCCAGCGGGGAGTGTTCTGCTGTCGTCAGATTACAGAGTCACGGGAGACACCGTTCTGTGT 1893
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   542 CTCCAGCGGGGAGTGTTCTGCTGTCGTCAGATTACAGAGTCACGGGAGACACCGTTCTGTGT  601

Qy  1894 TACTGCTGGGCCTGCCAGCTTCCGTGAGCTGAGCTGACCTATCAGTATCGGCAGAACATTCCT 1953
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   602 TACTGCTGGGCCTGCCAGCTTCCGTGAGCTGAGCTGACCTATCAGTATCGGCAGAACATTCCT  661

Qy  1954 GCTTCCGAGTT 1964
          |||||||||||
Db   662 GCTTCCGAGTT  672
```

```
Db   121 AGTCCACCACGCCATGAAATTCAATCATATCTGACACAGACAAGGTTCAAAAACTAAGC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2088 ATCCAGAGAGCCCTGAGCAGCTTTCAGCACTGGAGGTGAAGAGAGCCGTGTTTTTAAAATAC 2147
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ATCCAGAGAGCCCTGAGCAGCTTTCAGCACTGGAGGTGAAGAGAGCCGTGTTTTTAAAATAC  240

Qy  2148 AGAGCAAGCACGTCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGGACGCGCAGGGTCTCCG 2207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AGAGCAAGCACGTCAAGGTGTTTTCACAGCCCCCTGAGGGAAGGGACGCGCAGGGTCTCCG  300

Qy  2208 ACAGGTGTCTCGGGGTGACTCTTCTGTGGAGCTTTTACCCTCGAGTGAGTGAGACCCTCCCC 2267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACAGGTGTCTCGGGGTGACTCTTCTGTGGAGCTTTTACCCTCGAGTGAGTGAGACCCTCCCC  360

Qy  2268 AGAGCCCCGGGGGCGCAGCCCGCCCCTCGGTGAGCGCTGGGCAGGGCTCGTGGTGGCA 2327
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AGAGCCCCGGGGGCGCAGCCCGCCCCTCGGTGAGCGCTGGGCAGGGCTCGTGGTGGCA  420

Qy  2328 TCAGCAGCAGACGAAGCCTTTCTGTAACATGCGGCCGTCCCGCCCAGAGAGGGGCAGTTT 2387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TCAGCAGCAGACGAAGCCTTTCTGTAACATGCGGCCGTCCCGCCCAGAGAGGGGCAGTTT  480

Qy  2388 TGCTCTTTTGTACATTTTCCGAAACTACACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAGT 2447
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TGCTCTTTTGTACATTTTCCGAAACTACACAGTTAAAGCAGAAGTCTGTTTTCAGGAAAAGT  540

Qy  2448 TTCAAGGGAAGGGCAAGTTTATCAAAAACATTGTTCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTT 2507
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TTCAAGGGAAGGGCAAGTTTATCAAAAACATTGTTCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTT  600

Qy  2508 ACAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACAG 2554
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ACAGCCTACAGGACGTACACAATATCCTGCTGCTGGGAAAACCACAG  647
```

RESULT 6

BG395714

LOCUS   BG395714   840 bp   mRNA   linear   EST 12-MAR-2001
DEFINITION 6024588365F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4580633 5',
            mRNA sequence.
ACCESSION   BG395714
VERSION   BG395714.1 GI:13289162
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 840)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1300 row: a column: 18
            High quality sequence stop: 775.
FEATURES   Location/Qualifiers
     source   1..840
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4580633"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_16"
                     /note="Organ: eye; Vector: pOTB7; Site 1: XhoI; Site 2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'

RESULT 7

AU134937

LOCUS   AU134937   680 bp   mRNA   linear   EST 02-AUG-2002
DEFINITION AU134937 PLACE1 Homo sapiens cDNA clone PLACE1000852 5', mRNA
            sequence.
ACCESSION   AU134937
VERSION   AU134937.1 GI:11095476
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 680)

AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
           Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
           Isogai,T.
TITLE   HRI human cDNA project
JOURNAL   Unpublished (2000)
COMMENT   Contact: Takao Isogai
           Genomics Laboratory
           Helix Research Institute
           1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
           Tel: 81-438-52-3975
           Fax: 81-438-52-3986
           Email: genomics@hri.co.jp
           HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
           Research Institute; cDNA library construction: Department of
           Virology, Institute of Medical Science, University of Tokyo, and
           Helix Research Institute.
FEATURES   Location/Qualifiers
     source   1..680
                 /organism="Homo sapiens"
                 /mol_type="mRNA"
                 /db_xref="taxon:9606"
                 /clone="PLACB1000852"
                 /tissue_type="placenta"
                 /clone_lib="PLACB1"
                 /note="Vector: pME18SFL3"
ORIGIN

Query Match   22.4%; Score 600; DB 9; Length 680;
Best Local Similarity 100.0%; Pred. No. 1.4e-110;
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1503 CCAGGATTACGTGTGCCCTCTGTGAAGGAGCCACGCCCTGTGCACCTGCTTCCAGCC 1562
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CCAGGATTACGTGTGCCCTCTGTGAAGGAGCCACGCCCTGTGCACCTGCTTCCAGCC 60

Qy 1563 CATGCCCGACCGGAGAGCGGAGAGCCGGAGGACCCGTGTCGCCCCTCAGCAGTGTGC 1622
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CATGCCCGACCGGAGAGCGGAGAGCCGGAGGACCCGTGTCGCCCCTCAGCAGTGTGC 120

Qy 1623 GGTCTGCCTGCAGCCTTTCTGCCACCTGTACTGGGGCTGCACCCGGACCGGCTGCTACGG 1682
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 GGTCTGCCTGCAGCCTTTCTGCCACCTGTACTGGGGCTGCACCCGGACCGGCTGCTACGG 180

Qy 1683 CTGCCTGGCCCCGTTTGTGAGCTCAACCTGGGTGGTGACAAGTGTCTGGACGGCGTGCTGA 1742
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 CTGCCTGGCCCCGTTTGTGAGCTCAACCTGGGTGGTGACAAGTGTCTGGACGGCGTGCTGA 240

Qy 1743 CAACAGCTACGAGTCAGATCTGAAGAATTACCTGGCAACCAGAGGTTTGACATG 1802
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 CAACAGCTACGAGTCAGATCTGAAGAATTACCTGGCAACCAGAGGTTTGACATG 300

Qy 1803 GAAAAACATGTTGACCGAGAGCCTTCGTGCGCTCTCCAGCGGGGAGTGTTTCTGCTGTCTGA 1862
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GAAAAACATGTTGACCGAGAGCCCTCGTGTGGCTCTCCAGCGGGGAGTGTTTCTGCTGTCTGA 360

Qy 1863 TTACAGAGTCACGGGAGACACCCGTTCTGTGTTACTGCTGTGTACTGGGCTGCGGCAGCTTCCGTGA 1922
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 TTACAGAGTCACGGGAGACACCCGTTCTGTGTTACTGCTGTGTGACTGGGCTGCGGCAGCTTCCGTGA 420

Qy 1923 GCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATC 1982
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 GCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATC 480

Qy 1983 CCGTCCTGACTGCTACTGGGGCCCGTAACTGCCCGCACTCCAGGTGAAAGCTCACCACGCCAT 2042
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 CCGTCCTGACTGCTACTGGGGCCGTAACTGCCCGCACTCCAGGTGAAAGCTCACCACGCCAT 540

Qy 2043 GAAATTCAATCATATCTGTGAACAGACAAGTTCAAAAACTAAGCATCCAGAGCATCCAGAGCCCTGA 2102
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 GAAATTCAATCATATCTGTGAACAGACAAGTTCAAAAACTAAGCATCCAGAGCATCCAGAGCCCTGA 600

RESULT 8

Qy   949 GACGTCAGACGAGCGGCTGGGAAGCCAGACAAGATGAGGAGGACGCTGACATGCATCATC 1008
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   395 GACGTCAGACGAGCGGCTGGGAAGCCAGACAAGATGAGGAGGACGCTGACATGCATCATC 454

Qy   1009 TGCCAGGACCTGCTGCACGACTGCGTGAGTGTTGCAGCCTGAGTTTGCAGCCTTCTGCCG 1068
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   455 TGCCAGGACCTGCTGCACGACTGCGTGAGTGTTGCAGCCTGAGTTTGCAGCCTTCTGCGGG 514

Qy   1069 GCTTGCTACTCGGGCTGGATGGAGGCGCTGCTCCCTGTGTCCTACCTGCCGCGTGTCCCGTG 1128
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   515 GCTTGCTACTCGGGCTGGATGGAGGCGCTGCTCCCTGTGTCCTACCTGCCGCGTGTCCCGTG 574

Qy   1129 GAGCGGATCGTAAAAACCACATCCTCAACAACCTCGTGGAAGCATACCTCATCCAGCAT 1188
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   575 GAGGGGATCGTAAAAACCACATCCTCAACAACCTCGTGGAAGCATACCTCATCCAGCAT 634

Qy   1189 CCAGACAAGAGTGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACT 1248
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   635 CCAGACAAGAGTGCAGTGAAGAAGATGTGCAAAGTATGGATGCCAGGAATAAAATCACT 694

Qy   1249 CAAGACAATGCTGCAGCCCAAAGTCA 1273
           |||||||||||||||||||||||||
Db   695 CAAGACAATGCTGCAGCCCAAAGTCA 719

RESULT 10
BM542545
LOCUS   BM542545   1143 bp   mRNA   linear   EST 20-FEB-2002
DEFINITION   AGENCOURT_6426028 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5520909
             5', mRNA sequence.
ACCESSION   BM542545
VERSION   BM542545.1 GI:18772149
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1143)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM12185 row: k column: 22
             High quality sequence stop: 666.

FEATURES   Location/Qualifiers
     source   1..1143
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5520909"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.1 kb. "

ORIGIN

Query Match   21.0%; Score 563; DB 12; Length 1143;
Best Local Similarity 100.0%; Pred. No. 2e-103;
Matches 563; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   2099 CTGAGCAGCTTTCAGCACTGGAGGTGAAGAGAGCGTGTTTTTAAAATACAGAGACAAGCA 2158
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   1336 AGTGAGTCCTCAGACATTAGCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGA 1395
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AGTGAGTCCTCAGACATTAGCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGA 480

Qy   1396 AGGCAGGCGGCGCAGCCTCCCCACTGCCCAGCACCCGAGGGGCGAGCCAGGAGCCCCACAG 1455
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AGGCAGGCGGCGCAGCCTCCCCACTGCCCAGCACCCGAGGGGCGAGCCAGGAGCCCCACAG 540

Qy   1456 GCCCTGGGGGATGCACCCTCCAGTCCGTCAGCCTGACGACAG 1498
           |||||||||||||||||||||||||||||||||||||||||
Db   541 GCCCTGGGGGATGCACCCTCCAGTCCGTCAGCCTGACGACAG 583

RESULT 9
BM459647
LOCUS   BM459647   916 bp   mRNA   linear   EST 05-FEB-2002
DEFINITION   AGENCOURT_6418058 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5534277
             5', mRNA sequence.
ACCESSION   BM459647
VERSION   BM459647.1 GI:18508687
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 916)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM12220 row: h column: 22
             High quality sequence stop: 642.

FEATURES   Location/Qualifiers
     source   1..916
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5534277"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.1 kb. "

ORIGIN

Query Match   21.1%; Score 565; DB 12; Length 916;
Best Local Similarity 100.0%; Pred. No. 9.7e-104;
Matches 565; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   709 GGGTCTGGGGGTGGTGGCATCTCCCTAAAGGAAGTGGTCCCTCGTGGCAAGTGATGAA 768
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   155 GGGTCTGGGGGTGGTGGCATCTCCCTAAAGGAAGTGGTCCCTCGTGGCAAGTGATGAA 214

Qy   769 GTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAGACTGCGTCCTTTCGTCTTGCGTTGGAA 828
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   215 GTCTCCAGCTTTGCCTCAGCTCTCCCAGACAGAAGACTGCGTCCTTTCGTCTTGCGTTGGAA 274

Qy   829 CCCCAGGATCAGGAGGATTCGGAGCCCGTGAAGAAAGAAAATGAGAGAGGAGATGGGGACCTT 888
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   275 CCCCAGGATCAGGAGGATTCGGAGCCCGTGAAGAAAGAAAATGAGAGAGGAGATGGGGACCTT 334

Qy   889 GACCTGAACGGGCAGTTGTTGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAG 948
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   335 GACCTGAACGGGCAGTTGTTGTCGCACAACCGCGTAGAAATGCCCAAACCGTCCACGAG 394

/clone_lib="NIH_MGC_21"
/note="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

```
ORIGIN

Query Match          20.9%;   Score 560;  DB 10;   Length 666;
Best Local Similarity 99.8%;   Pred. No. 1.3e-102;
Matches 610;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;

Qy  1156 AACAACCTGTGGAAGCATACCTCATCCAGCATCCAGCATCCAGACAAGAGTGCAGTGAAGAAGAT 1215
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 AACAACCTGTGGAAGCATACCTCATCCAGCATCCAGCATCCAGACAAGAGTGCAGTGAAGAAGAT 60

Qy  1216 GTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGG 1275
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GTGCAAAGTATGGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGG 120

Qy  1276 CGGTCTTTTCTGATGAAGAAGGGAGTTCAGAGAGCCTGCTGCTGGAGCGTGTCAGACGTTGAC 1335
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CGGTCTTTTCTGATGAAGAAGGGAGTTCAGAGAGCCTGCTGCTGGAGCGTGTCAGACGTTGAC 180

Qy  1336 AGTGAGTCCTCAGACATTAGCCAGCCATACGTTGTGCCGCAGTGTCCTGAGTACAGA 1395
         ||||||||||||||||||||||||||||||||| |||||||||||||||||||| |
Db   181 AGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGCCGCAGTGTCCTGAGTACACAGA 240

Qy  1396 AGGCAGGCGGCGGCGCAGCCTCCCCACTGCGCCCAGCACCCGAGGGCGGAGCCCCACAG 1455
         ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   241 AGGCAGGCGGCGGCGCAGCCTCCCCACTGCCCCAGCACCCGAGGGCGGAGCCCCACAG 300

Qy  1456 GCCCTGGGGATGCACCCTCCACGTCCGTCAGCCTGACGACAGCAGTCCAGATTACGTG 1515
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GCCCTGGGGATGCACCCTCCACGTCCGTCAGCCTGACGACAGCAGTCCAGATTACGTG 360

Qy  1516 TGCCCTCTGCAAGGAAGCCACCGCCACCGCCCGTGCTTCCAGCCCATGCCCGACCGG 1575
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TGCCCTCTGCAAGGAAGCCACCGCCACCGCCCGTGCTTCCAGCCCATGCCCGACCGG 420

Qy  1576 AGAGCGGAGCGCGAGCAGGACAGGACCCGCGTGTCGCCCCTCCAGCGTGTGCGGTCTCGCCTGCAG 1635
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AGAGTGGAGCGCGAGCAGGACAGGACCCGCGTGTCGCCCCTCCAGCGTGTGCGGTCTCGCCTGCAG 480

Qy  1636 CCTTTCTGCCACCTGTACTGGGGCTGCACCCGGACCCGGGCTGCTCACGGCTGCTGGCCCCG 1695
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CCTTTCTGCCACCTGTACTGGGGCTGCACCCGGACCCGGGCTGCTCACGGCTGCTGGCCCCG 540

Qy  1696 TTTTGTGAGTCTCAACCTGGGTGACAAGTGTCTGGACGGTGCTGGACCAACAACACAGCTAC 1755
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TTTTGTGAGTCTCAACCTGGGTGACAAGTGTCTGGACGGTGCTGGACCAACAACACAGCTAC 600

Qy  1756 GAGTCAGACAT 1766
         |||||||||||
Db   601 GAGTCAGACAT 611
```

RESULT 12
AL539797/c
LOCUS   AL539797   1201 bp   mRNA   linear   EST 31-MAY-2003
DEFINITION AL539797 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF031YA23 3-PRIME, mRNA sequence.
ACCESSION   AL539797
VERSION   AL539797.2 GI:31264360
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1201)

```
Db    37 CTGAGCAGCTTCAGCACTGGAGTGAAGAGAGCGGTGTGTTTTTAAAATACAGAGACAAGCA 96
Qy  2159 CGTCAAGGTGTTTTCACAGCCCCCTGAGGGAGGGACGCAGGGTCTCCGACAGGTGCTCT 2218
         ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    97 CGTCAAGGTGTTTTCACAGCCCCCTGAGGGAGGGACCCAGGGTCTCCGACAGGTGCTCT 156

Qy  2219 GGGGTGACTCTTCTGTGGAGCTTTTACCCTCGAGTGAGACCCTCCCAGAGCCCCGGG 2278
         |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   157 GGGGTGACTCTTCTGTGGAGCTTTTACCCTCTGAGTGAGACCCTCCCAGAGCCCCGGG 216

Qy  2279 GGCCGCAGCCGCGCCCTCCTGGTGAGCGGCTCGGCCAGGACTCGTTGGTGGCATCAGCACCAGA 2338
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   217 GGCCGCAGCCGCGCCCTCCTGGTGAGCGGCTCGGCCAGGACTCGTTGGTGGCATCAGCAGCAGA 276

Qy  2339 GACGAAGCCTTTCTGTAACATGCGGCCGTCCCCGCGAGAGGGGCAGTTTGCTCTTTTGT 2398
         |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   277 GACGAAGCCTTTCTGTAACATGCGGCCGTCCCCGCGAGAGGGGCAGTTTTGCTCTTTTGT 336

Qy  2399 ACATTTTCCGAAACTACAGTTAAAGCAGAGAGTCTGTTTTCAGGAAAAAGTTTCAAGGGGAGA 2458
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   337 ACATTTTCCGAAACTACAGTTAAAGCAGAGAGTCTGTTTTCAGGAAAAAGTTTCAAGGGGAGA 396

Qy  2459 AGGGCAAGTTTATCAAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTACAG 2518
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   397 AGGGCAAGTTTATCAAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTTACAGCCTACAG 456

Qy  2519 GACGTACACAAATATCCTGCTCGTGTCGGGAAAAACCACACAGCATTTTATCTATTTTATTTTAA 2578
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   457 GACGTACACAAATATCCTGCTCGTGTCGGGAAAAACCACACAGCATTTTATCTATTTTATTTTAA 516

Qy  2579 TAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATA 2638
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   517 TAGGTTTGGTGCTTATCTTCTAATAAGATTTAAATGTCACAAACTGTAGCACAAATAATA 576

Qy  2639 TAATTTATATATTTACAAAATTGAC 2661
         |||||||||||||||||||||||||
Db   577 TAATTTATATATTTACAAAATTGAC 599
```

RESULT 11
BE407732
LOCUS   BE407732   666 bp   mRNA   linear   EST 21-JUL-2000
DEFINITION 601299777F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3629682 5',
            mRNA sequence.
ACCESSION   BE407732
VERSION   BE407732.1 GI:9344182
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 666)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM316 row: j column: 19
            High quality sequence stop: 640.
FEATURES   Location/Qualifiers
     source   1..666
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3629682"
                     /clone_lib="choriocarcinoma"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"

Left column

```
AUTHORS     Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:12865347.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6792.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF031AA12NP1&cluster=6792.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF031AA12NP1.
FEATURES             Location/Qualifiers
     source          1..1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF031YA23"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

Query Match        20.1%;  Score 539;  DB 9;  Length 1201;
Best Local Similarity  100.0%;  Pred. No. 1.1e-98;
Matches  539;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy   1704 GCTCAACCTGGGTGACAAGTGTCTGACGCGCGTGCTGAACAACAACAGCTACCAGTCAGA 1763
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    940 GCTCAACCTGGGTGACAAGTGTCGACGCGGTGCTGAACAACAACAGCTACCAGTCAGA 881

Qy   1764 CATCCTGAAGAATTACCTGGCAACCAGAGGTTGACATGGAAAAACATGTTGACCGAGAG 1823
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    880 CATCCTGAAGAATTACCTGGCAACCAGAGGTTTGACATGGAAAAACATGTTGACCGAGAG 821

Qy   1824 CCTCGTGGCTCTCCAGCGGCGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACAC 1883
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    820 CCTCGTGGCTCTCCAGCGGCGGGAGTGTTTCTGCTGTCTGATTACAGAGTCACGGGAGACAC 761

Qy   1884 CGTTCTGTGTTACTGCTGTGGCGCTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCGGCA 1943
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    760 CGTTCTGTGTTACTGCTGTGGCGCTGGCCTGCGCAGCTTCCGTGAGCTGACCTATCGGCA 701

Qy   1944 GAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCGTCCTGACTGCTACTGGGG 2003
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    700 GAACATTCCTGCTTCCGAGTTGCCAGTGGCCGTAACATCCGTCCTGACTGCTACTGGGG 641

Qy   2004 CCGTAACTGCGCGCACTCAGGTGAAAGCTCACCACGCCACCAGCGCCCATGAAATTCAATCATATATCTGTGA 2063
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    640 CCGTAACTGCGCGCACTCAGGTGAAAGCTGAAGCTCACCACGCCACCAGCGCCCATGAAATTCAATCATATATCTGTGA 581

Qy   2064 ACAGACAAGGTTCAAAAACTAAGCATCCAGAGAGCCCTGAGCCAGCTTTCAGCACTGGAGGT 2123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    580 ACAGACAAGGTTCAAAAACTAAGCATCCAGAGAGCCCTGAGCCAGCTTTCAGCACTGGAGGT 521

Qy   2124 GAAGAGAGCCGTGTTTTTAAAATACAGAGACAAGCAGCACGTCAAGGTGTTTTCAGCAGCCCCCT 2183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    520 GAAGAGAGCCGTGTTTTTAAAATACAGAGACAAGCAGCACGTCAAGGTGTTTTCAGCAGCCCCT 461

Qy   2184 GAGGGAAGGGACGCGACAGGTGCTCCGACAGGTGCTCTTCGTGTGGGAGCTTT 2242
          ||||||||||||||||||||||||||||||||||||||||||||
Db    460 GAGGGAAGGGACGCGACAGGGTGCTCCGACAGGTGCTCCTTCGTGTGGGAGCTTT 402
```

Right column

```
RESULT 13
BF971838
LOCUS       BF971838   602240424F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4129076 5',  934 bp   mRNA   linear   EST 22-JAN-2001
DEFINITION  602240424F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4129076 5',
            mRNA sequence.
ACCESSION   BF971838
VERSION     BF971838.1  GI:12339053
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 934)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1189 row: p column: 05
            High quality sequence stop: 748.
FEATURES             Location/Qualifiers
     source          1..934
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4129076"
                     /tissue_type="leiomyosarcoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_46"
                     /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN

Query Match         19.4%;  Score 519;  DB 10;  Length 934;
Best Local Similarity  99.8%;  Pred. No. 1.4e-94;
Matches  639;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  1;

Qy   1908 GGGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCC 1967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGCAGCTTCCGTGAGCTGACCTATCAGTATCGGCAGAACATTCCTGCTTCCGAGTTGCC 60

Qy   1968 AGTGGCCTAACATCCGTCCTGACTGCTACTGGGGCCGTAACTGCTGCCCACTCAGGTGAA 2027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGTGGCCTAACATCCCGTCCTGACTGCTACTGGGGCCGTAACTGCTGCCCACTCAGGTGAA 120

Qy   2028 AGCTCACCACGCCATGAAATTCAATCATATATCTGTGAACAGACAAGGTTCAAAAACTAAGC 2087
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGCTCACCACGCCATGAAATTCAATCATATATCTGTGAACAGACAAGGTTCAAAAACTAAGC 180

Qy   2088 ATCCAGAGCCCTGAGCCAGCTTTCAGCACTGGAGGTGAAGAGAGCCGTGTTTTTAAAATAC 2147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATCCAGAGCCCTGAGCCAGCTTTCAGCACTGGAGGTGAAGAGAGCCGTGTTTTTAAAATAC 240

Qy   2148 AGAGACAAGCACGTCAAGGTGTTTTCAGCAGCCCCCTGAGGGGAAGGGACGCAGCAGGGTCTCCG 2207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGAGACAAGCACGTCAAGGTGTTTTCAGCAGCCCCCTGAGGGGAAGGGACGCAGCAGGGTCTCCG 300

Qy   2208 ACAGGTGCTCTCGGGTGACTCTTCGTGGAGCTTTTTACCCTCGTGGAGCTCCTGAGTGAGCCCCCCC 2267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    13 CCGATCCCCGGGTCCCTCCGTCGTCGCCGCCACTCAGGTGTGCTTTGAGGAGGAACCACAGC  72
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  2268 AGAGCCCCGGGGGGGGGGCCGCAGCCCGCCCCTCCTGTGAGCGGCTGGGCAGGGCTGTGGGCA 2327
         |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||||
Db   361 AGAGCCCCGGGGGGGGGGCCGCAGCCCG--CCTCCTGTGATGAGCGCTGCTGGGCAGGGCTGTGGTGGCA  419

Qy  2328 TCAGCAGCAGAGACGAAGCCTTTCTGTAACATGCGGCCGTCCCGCCGAGAGGGGCAGTTT 2387
         |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db   420 TCAGCAGCAGAGACGAAGCCTTTCTGTAACATGCGGCCGTCCCGCCGAGAGGGGCAGTTT  479

Qy  2388 TGCTCTTTTGTACATTTCCGAAACTACAGTAAAGCAGAGAGTCTGTTTTCAGGAAAAGT 2447
         |||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||
Db   480 TGCTCTTTTGTACATTTCCGAAACTACAGTAAAGCAGAGAGTCTGTTTTCAGGAAAAGT  539

Qy  2448 TTCAAGGGAGAAGGGCCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTT 2507
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   540 TTCAAGGGAGAAGGGCCAAGTTTATCAAAAACATTGTTTCAGGAGAAGGGAGCATAAGTTT  599

Qy  2508 ACAGCCTACAGGACGTACACAATATCCTGCTGGGAAA 2547
         |||||||||||||||||||||||||||||||||||||
Db   600 ACAGCCTACAGGACGTACACAATATCCTGCTGGGAAA  639
```

RESULT 14
CA916737
LOCUS   CA916737   729 bp   mRNA   linear   EST 14-MAR-2003
DEFINITION   30LSPRIM2-5A human lymphocyte Matchmaker cDNA library Homo sapiens
               cDNA clone L5 5' similar to CHFR RF; cell cycle check point, mRNA
               sequence.
ACCESSION   CA916737
VERSION   CA916737.1 GI:28951974
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 729)
AUTHORS   Scarafia,L.E., Stoufer,S.D. and Swinney,D.C.
TITLE   Identification of Ring Finger proteins that interact with UbcH5a,
               an ubiquitin-conjugating enzyme
JOURNAL   Unpublished (2002)
COMMENT   Contact: Scarafia LE
               Enzymology
               Roche Bioscience
               3401 Hillview Ave, S3-1, Palo Alto, CA 94304-1397, USA
               Tel: 650 354 7997
               Fax: 650 354 7554
               Email: liliana.scarafia@roche.com
               Seq primer: matchmaker 5' AD (ctattcgatgatgaagataccccaccaaaccc)
               POLYA=No.
FEATURES   Location/Qualifiers
     source   1..729
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L5"
                     /cell_type="B cell"
                     /cell_line="EBV-transformed human peripheral blood
                     lymphocyte; B cell population; IG (+)"
                     /lab_host="yeast/E.coli"
                     /clone_lib="human lymphocyte Matchmaker cDNA library"
                     /note="Vector: pACT; Site 1: XhoI; Site 2: XhoI;
                     Matchmaker yeast two-hybrid sytem from Clontech;pACT has
                     GAL4 AD under ADH1 promoter. This oligo-dT primed library
                     was screened with human UbcH5a as bait, to obtain
                     interacting proteins."
ORIGIN

Query Match   19.1%; Score 513; DB 14; Length 729;
Best Local Similarity 99.5%; Pred. No. 2.6e-93;
Matches 663; Conservative   0; Mismatches   3; Indels   0; Gaps   0;

```
Qy   560 CCGATCCCCGGGTCCCTCCGTCGCCGCCGCCACTCAGGTGTGCTTTGAGGAACCACAGC  619
```

```
Db    13 CCGATCCCCGGGTCCCTCCGTCGTCGCCGCCACTCAGGTGTGCTTTGAGGAGGAACCACAGC  72
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   620 CATCAACATCGACGTCAGACCTCTTCCCACAGCCTCGGCCTCTTCCACGGAGCCTTCTC  679
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    73 CATCAACATCGACGTCAGACTCTTCCCACAGCCTCCCACAGCCTCGGCCTCTCCACGGAGCCTTCTC  132

Qy   680 CTGCAGCGGAGAGCGGTTCCTCCTGTGGGTGGTGGCATCTCCCCTAAAG  739
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   133 CTGCAGCGGAGAGCGGTTCCTCCTGTGGGTGGTGGCATCTCCCCTAAAG  192

Qy   740 GAAGTGGTCCCTCTGGCAAGTGATGAAGTCTCCAGCTTGCCTCAGCTCTCCCAGACA  799
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   193 GAAGTGGTCCCTCCTGTGGCAAGTGATGAGTGATGAAGTCTCCAGCTTGCCTCAGCTCTCCCAGACA  252

Qy   800 GAAAGACTGCGTCCTTTCGTCGTTGGAACCCCAGGATCAGGAGGATTGGAGCCCGTGA  859
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   253 GAAAGACTGCGTCCTTTGTCGTTGGAACCCCAGGATCAGGAGGATTTGGAGCCCGTGA  312

Qy   860 AGAGAAAATGAGAGAGAGTGGGGACCTTGACCTGACCTGTGTGGTCGCACAAC  919
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   313 AGAAGAAAATGAGAGAGAGTGGGGACCTTGACCTGACCTGTGTGGTCGCACAAC  372

Qy   920 CGCGTAGAAATGCCCAAACCGTCACGAGACGTCAGAGCAGCGTCAGAGCAGCAGACA  979
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   373 CGCGTAGAAATGCCCAAACCGTCCAAACCGTCCACGAGACGTCAGAGCAGCAGACA  432

Qy   980 AGAATGGAGGAGACCTGACATGCATCATCTGCCAGGACTGCGTGCGGTGAGTT  1039
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   433 AGAATGGAGGAGACCTGACATGCATCATCTGCCAGGAACTGCGTCACGACTGCGTGAGTT  492

Qy  1040 TGCAGCCCTGCATGGCACACGTTCTGCGCGGCTTGCTACTCGGCTGGATGGAGCGGCTCGT 1099
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   493 TGCAGCCCTGCATGGCACACGTTCTGCGCGGCTTGCTACTCCGGCTGGATGGAGCGGCTCGT  552

Qy  1100 CCCTGTGTCCTACCTGCCGCTGTCCGTGGAGCGGATCGTGTAAAAACCACCATCCTCAACA 1159
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   553 CCCTGTGTCCTACCTGCCGCTGTCCGTGGAGCGGATCGTGTAAAAACCACCATCCTCAACA  612

Qy  1160 ACCTCTGGAAGCATACCTCATCCAGCATCCAGCATCTGCCAGTGAAGAAGATGTGC 1219
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   613 ACCTCTGGAAGCATACCTCATCCAGCATCCAGCATCTGCCAGTGAAGAAGATGTGC  672

Qy  1220 AAAGTA 1225
         ||||||
Db   673 AAAGTA  678
```

RESULT 15
BE280165
LOCUS   BE280165   690 bp   mRNA   linear   EST 13-JUL-2000
DEFINITION   601158370F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3504928 5',
               mRNA sequence.
ACCESSION   BE280165
VERSION   BE280165.1 GI:9155072
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 690).
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Preparation: Ling Hong/Rubin Laboratory
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
               Plate: LLCM180 row: 1 column: 17

```
FEATURES      High quality sequence stop: 626.
   source      Location/Qualifiers
               1..690
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:3504928"
               /tissue_type="choriocarcinoma"
               /lab_host="DH10B (phage-resistant)"
               /clone_lib="NIH_MGC_21"
               /note="Organ: placenta; Vector: pOTB7; Site 1: XhoI;
               Site 2: EcoRI; cDNA made by oligo-dT priming;
               Directionally cloned into EcoRI/XhoI sites using the
               following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
               for average insert size 1.8kb.  Library constructed by
               Ling Hong in the laboratory of Gerald M. Rubin (University
               of California, Berkeley) using ZAP-cDNA synthesis kit
               (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Query Match          18.2%;  Score 488;  DB 10;  Length 690;
Best Local Similarity   99.8%;  Pred. No. 2.6e-88;
Matches  538;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   1156 AACAACCTCGTGGAAGCATACCTCATCCAGCATCCAGACAAGAGTCGCAGTGAAGAAGAT 1215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AACAACCTCGTGGAAGCATACCTCATCCAGCATCCAGACAAGAGTCGCAGTGAAGAAGAT   60

Qy   1216 GTGCAAAGTATGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGG 1275
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTGCAAAGTATGATGCCAGGAATAAAATCACTCAAGACATGCTGCAGCCCAAAGTCAGG  120

Qy   1276 CGGTCTTTTTCTGATGAAGAAGGGAGTTCAGAGACCTGCTCGGAGCGTGTCAGACGTTGAC 1335
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGGTCTTTTTCTGATGAAGAAGGGAGTTCAGAGACCTGCTCGGAGCGTGTCAGACGTTGAC  180

Qy   1336 AGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGA 1395
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGTGAGTCCTCAGACATTAGCCAGCCATACGTCGTGTGCCGGCAGTGTCCTGAGTACAGA  240

Qy   1396 AGGCAGGCGGCGGCAGCCTCCCCACTGCCCAGACACCCGAGGGGCGAGCCAGGAGCCCCACAG 1455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AGGCAGGCGGCGGCAGCCTTCCCCACTGCCCAGACACCCGAGGGGCGAGCCAGGAGCCCCACAG  300

Qy   1456 GCCCTGGGGATGCACCCTCCACGTCCGTCAGCCTGACAGACAGCAGTCCAGGATTACGTG 1515
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCCCTGGGGATGCACCCTCCACGTCCGTCAGCCTGACGACGACAGTCCAGGATTACGTG  360

Qy   1516 TGCCCTCTGCAAGGAAGCCACGCCCTGCACTTGCTTCCAGCCCATGCCCGACCGG 1575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGCCCTCTGCGAAGGAAGCCACGCCCTGCACTTGCTTCCAGCCCATGCCCGACCGG  420

Qy   1576 AGAGCGGAGCGGCGGAGCGAGGCGCAGGACCCGGTGTCGGCCCCTCAGCAGTGTGCGGTCTGCCTGCAG 1635
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGAGTGGAGCGGCGGCAGCAGGACCCGGTGTCGGCCCCTCAGCAGTGTGCGGTCTGCCTGCAG  480

Qy   1636 CCTTTCTGCCACCTGTACTGGGGCTGCACCCGACCGGCTGCTACGGCTGCTGGCCCC 1694
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCTTTCTGCCACCTGTACTGGGGCTGCACCCGACCGGCTGCTACGGCTGCTGGCCCC  539
```

Search completed: May 15, 2004, 07:58:44
Job time : 10042 secs

```
                    Copyright    GenCore version 5.1.6
                    Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:        May  7, 2004, 14:36:57 ; Search time 136.025 Seconds
                                        (without alignments)
                               1379.240 Million cell updates/sec

Title:         US-10-048-046-2
Perfect score: 3585
Sequence:      1 MERPEEGKQSPPPQPWGRLL.........VKAHHAMKFNHICEQTRFKN 664

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:        1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    A_Geneseq_29Jan04:*
              1: geneseqp1980s:*
              2: geneseqp1990s:*
              3: geneseqp2000s:*
              4: geneseqp2001s:*
              5: geneseqp2002s:*
              6: geneseqp2003as:*
              7: geneseqp2003bs:*
              8: geneseqp2004s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	3585	100.0	664	4	AAB83843	Aab83843 Amino aci
2	3585	100.0	664	4	AAB20219	Aab20219 Human Chf
3	3504	97.7	652	4	AAB93168	Aab93168 Human pro
4	3168.5	88.4	623	5	ABB97233	Abb97233 Novel hum
5	3162.5	88.2	623	4	AAB93182	Aab93182 Human pro
6	1954.5	54.5	426	4	AAU15856	Aau15856 Human nov
7	1954.5	54.5	426	6	ABU54925	Abu54925 Human nov
8	1364	38.0	269	4	AAO08972	Aao08972 Human pol
9	843.5	23.5	230	4	AAU16317	Aau16317 Human nov
10	843.5	23.5	230	6	ABU55386	Abu55386 Human pro
11	566	15.8	128	6	ADA54231	Ada54231 Human pro
12	474	13.2	92	4	AAU16156	Aau16156 Human nov
13	474	13.2	92	6	ABU55225	Abu55225 Human nov
14	432	12.1	99	4	AAU16157	Aau16157 Human nov
15	432	12.1	99	4	AAU16576	Aau16576 Human nov
16	432	12.1	99	6	ABU55645	Abu55645 Human nov
17	432	12.1	99	6	ABU55226	Abu55226 Human nov
18	355	9.9	426	3	AAG20353	Aag20353 Arabidops
19	355	9.9	426	3	AAG37659	Aag37659 Arabidops
20	355	9.9	453	3	AAG20352	Aag20352 Arabidops
21	355	9.9	453	3	AAG37658	Aag37658 Arabidops
22	354	9.9	350	3	AAG20354	Aag20354 Arabidops
23	354	9.9	350	3	AAG37660	Aag37660 Arabidops
24	203	5.7	485	4	AAB95017	Aab95017 Human pro
25	160.5	4.5	352	5	AAU80358	Aau80358 Human cel
26	158	4.4	650	7	ADC31511	Adc31511 Human nov
27	154.5	4.3	731	6	ABJ19398	Abj19398 Human int
28	152	4.2	245	2	AAW59134	Aaw59134 Mus muscu
29	151.5	4.2	502	5	ABP51312	Abp51312 Human MDD
30	146.5	4.1	220	4	AAB94801	Aab94801 Human pro
31	146.5	4.1	220	6	AAU80360	Aau80360 Human cel
32	146.5	4.1	284	6	ABU69600	Abu69600 Human NF-
33	146	4.1	303	6	ABU92033	Abu92033 Human pro
34	143	4.0	1601	4	ABB60248	Abb60248 Drosophil
35	142.5	4.0	499	5	AAO22598	Aao22598 499-mer S
36	142.5	4.0	1213	6	AAO22597	Aao22597 1213-mer
37	141.5	3.9	1208	4	AAB95292	Aab95292 Human pro
38	139	3.9	825	6	ABP71596	Abp71596 HSV-2 ICP
39	139	3.9	4675	5	ABP70085	Abp70085 Human NOV
40	139	3.9	4691	5	ABP70084	Abp70084 Human NOV
41	138.5	3.9	353	5	AAU80359	Aau80359 Mouse cel
42	138.5	3.9	533	7	ADC31279	Adc31279 Human nov
43	138	3.8	255	4	AAG73609	Aag73609 Human col
44	137	3.8	1215	6	ABU11763	Abu11763 Human MDD
45	135	3.8	67	2	AAR60624	Aar60624 Varicella

ALIGNMENTS

```
RESULT 1
AAB83843
ID   AAB83843 standard; protein; 664 AA.
XX
AC   AAB83843;
XX
DT   22-AUG-2001  (first entry)
XX
DE   Amino acid sequence of a human ring finger protein designated FHAR1.
XX
KW   FHAR1; RING finger protein; cancer; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200142430-A1.
XX
PD   14-JUN-2001.
XX
PF   07-DEC-2000; 2000WO-US033094.
XX
PR   08-DEC-1999;  99US-00456876.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Zhou B, Zhu Y, Chaturvedi P, Hurle MR, Li X;
XX
DR   WPI; 2001-381663/40.
DR   N-PSDB; AAF89709.
XX
PT   New FHAR1 polypeptide, a member of the RING finger protein family for
PT   diagnosing and treating cancer, and for use in anti-cancer vaccines.
XX
PS   Claim 1; Page 19; 28pp; English.
XX
CC   The present sequence represents a FHAR1 polypeptide, which is a member of
CC   the RING finger protein family. FHAR1 is useful in the treatment of
CC   cancer, and as a vaccine for inducing an immunological response in a
CC   mammal. FHAR1 polynucleotides may also be used as a diagnostic reagent
CC   through detection of mutations in the associated gene, and for chromosome
CC   localization studies, and tissue expression studies. FHAR1 antibodies are
CC   useful to isolate and to identify clones expressing the polypeptides, or
CC   to purify the polypeptides by affinity chromatography and to treat cancer
XX
SQ   Sequence 664 AA;

Query Match             100.0%;  Score 3585;  DB 4;  Length 664;
Best Local Similarity   100.0%;  Pred. No. 9.8e-270;
Matches  664;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
FT   Misc-difference  580
FT                    /note= "Met in U2OS cells"
XX
PN   WO200109150-A2.
XX
PD   08-FEB-2001.
XX
PF   14-JUN-2000;  2000WO-US016391.
XX
PR   29-JUL-1999;  99US-0146194P.
XX
PA   (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI   Halazonetis T, Scolnick D;
XX
DR   WPI; 2001-182927/18.
DR   N-PSDB; AAF30352.
XX
PT   Novel nucleic acid sequence of mitotic checkpoint gene encoding a
PT   checkpoint with forkhead-associated domain and ring finger protein, for
PT   diagnosing tumourigenic cells and in screening for anticancer drugs.
XX
PS   Claim 8(a); Fig 4A-C; 85pp; English.
XX
CC   The present sequence is that of human mitotic checkpoint protein Chfr,
CC   having a forkhead associated domain (FHA) and a ring finger domain. The
CC   protein is required for regulation of the transition of cells from
CC   prophase to metaphase during mitosis. It has ubiquitin-protein ligase
CC   activity. The Chfr checkpoint was evident in primary human cells, but was
CC   inactivated in 4 of 8 human cancer cell lines. In U2OS cells, a mutation
CC   was identified that caused a Val to Met amino acid substitution in the
CC   highly conserved C-terminal Cys-rich region of the Chfr protein. In the
CC   absence of the Chfr checkpoint, cells subjected to mitotic stress
CC   condensed their chromosomes despite failing to separate their
CC   chromosomes. Chfr may monitor centrosome separation. Inactivation of the
CC   chfr gene (see AAF30352) in human cancer is theorized to underlie the
CC   increased sensitivity of cancer cells to antimitotic drugs. Polypeptides
CC   comprising the present sequence, or sequences comprising at least amino
CC   acids 31-103, 303-346 and/or 476-641 of this sequence, are claimed.
CC   Claimed methods of determining the tumourigenic potential of a cell
CC   comprise examining the cell for the presence of Chfr expression or for
CC   the presence of Chfr-mediated ubiquitin-protein ligase activity (in both
CC   cases, absence of expression indicating predisposition to tumourigenesis
CC   upon exposure to mitotic stress). A diagnostic kit for detecting the
CC   tumourigenic potential of cell cells comprises may comprise a ligand that
CC   binds to Chfr, such as an antibody or its fragment. Inhibitors of Chfr
CC   are identified by monitoring their effect on Chfr expression, and are
CC   used to retard the growth of cancer cells
XX
SQ   Sequence 664 AA;

Query Match             100.0%;  Score 3585;  DB 4;  Length 664;
Best Local Similarity   100.0%;  Pred. No. 9.8e-270;
Matches   664;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1  MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLJVSGDH   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLJVSGDH   60

Qy   61  CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY  120

Qy  121  ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS  180

Qy  181  TSDLFPTASASSTEPSPAGRERSSSCGSGGGISPKGSGPSVASDEVSSFASALPDRKTA  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TSDLFPTASASSTEPSPAGRERSSSCGSGGGISPKGSGPSVASDEVSSFASALPDRKTA  240

Qy  241  SFSSLEPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE  300
```

```
Qy    1  MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLJVSGDH   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLJVSGDH   60

Qy   61  CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY  120

Qy  121  ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS  180

Qy  181  TSDLFPTASASSTEPSPAGRERSSSCGSGGGISPKGSGPSVASDEVSSFASALPDRKTA  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TSDLFPTASASSTEPSPAGRERSSSCGSGGGISPKGSGPSVASDEVSSFASALPDRKTA  240

Qy  241  SFSSLEPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  SFSSLEPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE  300

Qy  301  TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE  360

Qy  361  AYLIQHPDKSRSEBDVQSMDARNKITQMLQPKVRRSFSDEEGSSEDLLELSDVDSESSD  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AYLIQHPDKSRSEBDVQSMDARNKITQMLQPKVRRSFSDEEGSSEDLLELSDVDSESSD  420

Qy  421  ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG  480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG  480

Qy  481  SHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCRTGCYGCLAPFCELN  540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  SHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCRTGCYGCLAPFCELN  540

Qy  541  LGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTRSLVALQRGVFLLSDYRVTGDTVL  600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  LGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTRSLVALQRGVFLLSDYRVTGDTVL  600

Qy  601  CYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCWGRNCRTQVKAHHAMKFNHICEQT  660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  CYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCWGRNCRTQVKAHHAMKFNHICEQT  660

Qy  661  RFKN  664
         ||||
Db  661  RFKN  664
```

RESULT 2
AAB20219

```
ID   AAB20219 standard; protein; 664 AA.
XX
AC   AAB20219;
XX
DT   14-MAY-2001  (first entry)
XX
DE   Human Chfr (checkpoint with FHA and ring finger) protein.
XX
KW   Checkpoint with forkhead associated domain and ring finger; Chfr; human;
KW   mitosis; cell cycle; tumour; diagnosis; antitumour; drug screening;
KW   ubiquitin-protein ligase.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   Domain         31..103
FT                  /label= Forkhead-associated_domain
FT   Domain         303..346
FT                  /label= Ring_finger-domain
FT   Region         476..641
FT                  /note= "cysteine-rich region"
```

```
Db    241 SFSSLEPDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE 300
Qy    301 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE 360
Qy    361 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEGGSSEDLLELSDVDSBSSD 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEGGSSEDLLELSDVDSBSSD 420
Qy    421 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG 480
Qy    481 SHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELN 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 SHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELN 540
Qy    541 LGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVL 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 LGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVL 600
Qy    601 CYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQT 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQT 660
Qy    661 RFKN 664
          ||||
Db    661 RFKN 664
```

```
RESULT 3
AAB93168
ID   AAB93168 standard; protein; 652 AA.
XX
AC   AAB93168;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:12100.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000;   2000JP-00118776.
PR   02-MAY-2000;   2000JP-00183767.
PR   09-JUN-2000;   2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 12100; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
```

```
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 652 AA;

Query Match             97.7%;  Score 3504;  DB 4;  Length 652;
Best Local Similarity   98.2%;  Pred. No. 1.9e-263;
Matches  652;  Conservative   0;  Mismatches   0;  Indels  12;  Gaps   1;
```

```
Qy      1 MERPBEGKQSPPPQPWGRLLRLGAEBGEPHVLLRKKREWTIGRRRGCDLSFPSNKLVSGDH 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MERPBEGKQSPPPQPWGRLLRLGAEBGEPHVLLRKKREWTIGRRRGCDLSFPSNKLVSGDH 60

Qy     61 CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY 120

Qy    121 ESLSEKQGMTQBSFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS 180
          |||||||||||||             |||||||||||||||||||||||||||
Db    121 ESLSEKQGMTQBSF--------DTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS 168

Qy    181 TSDLFPTASASTEPSPAGRERSSSCCSGGGGISPKGSGPSVASDEVSSFASALPDRKTA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    169 TSDLFPTASASTEPSPAGRERSSSCCSGGGGISPKGSGPSVASDEVSSFASALPDRKTA 228

Qy    241 SFSSLEPDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    229 SFSSLEPDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE 288

Qy    301 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    289 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE 348

Qy    361 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEGGSSEDLLELSDVDSBSSD 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    349 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEGGSSEDLLELSDVDSBSSD 408

Qy    421 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    409 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG 468

Qy    481 SHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELN 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    469 SHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELN 528

Qy    541 LGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVL 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    529 LGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVL 588

Qy    601 CYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQT 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    589 CYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQT 648

Qy    661 RFKN 664
          ||||
```

Db   649 RFKN 652

RESULT 4
ABB97233
ID   ABB97233 standard; protein; 623 AA.
XX
AC   ABB97233;
XX
DT   27-JUN-2002 (first entry)
XX
DE   Novel human protein SEQ ID NO: 501.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US026015.
XX
PR   11-SEP-2000; 2000US-00659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   N-PSDB; ABN32419.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis.
XX
PS   Example 2; SEQ ID NO 501; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC   and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC   rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a protein of the invention
XX
SQ   Sequence 623 AA;

Query Match   88.4%; Score 3168.5; DB 5; Length 623;
Best Local Similarity   88.7%; Pred. No. 2.2e-237;
Matches 595; Conservative   5; Mismatches 16; Indels   55; Gaps   2;

Qy   1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   60

Qy   61 CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY 120

Qy   121 ESLSEKQGMTQESFEANKENVFHGTKDTSGAGARGCADPRVPPSSPATQVCFEPQPSTS 180
           ||||||||||||||||||| :
Db   121 ESLSEKQGMTQESFE-- 135

Qy   181 TSDLFPTASA-------SSTEPSPAGRERSSSCCGSGGGGISPKGSGSPSVASDEVSSFASA 233
           |||||| | ::| |||||||||||||||||||||||||||
Db   136 ---MVPCCVAQAGLKLLGSSDPPTLASQSIVITGSGGGGISPKGSGSPSVASDEVSSFASA 192

RESULT 5
AAB93182
ID   AAB93182 standard; protein; 623 AA.
XX
AC   AAB93182;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human protein sequence SEQ ID NO:12128.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000;   2000JP-00118776.
PR   02-MAY-2000;   2000JP-00183767.
PR   09-JUN-2000;   2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 12128; 2537pp + Sequence Listing; English.

Qy   234 LPDRKTASFSSLEPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAG 293
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   193 LPDRKTASFSSLEPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAG 252

Qy   294 KPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNH 353
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   253 KPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNH 312

Qy   354 ILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSD 413
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   313 ILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSD 372

Qy   414 VDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQD 473
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   373 VDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPPTSVSLTTAVQD 432

Qy   474 YVCPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCL 533
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   433 YVCPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCL 492

Qy   534 APFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYR 593
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   493 APFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYR 552

Qy   594 VTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKF 653
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   553 VTGDTVLCYCCGLRSFRELTYQYQQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKF 612

Qy   654 NHICEQTRFKN 664
           |||||||||||
Db   613 NHICEQTRFKN 623

Query Match   88.2%;   Score 3162.5; DB 4;   Length 623;
Best Local Similarity   88.5%;   Pred. No. 6.5e-237;
Matches 594;   Conservative   5;   Mismatches   17;   Indels   55;   Gaps   2;

```
Qy    1    MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1    MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   60

Qy    61   CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61   CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY  120

Qy    121  ESLSEKQGMTQESFEANKENVFHGTKDTSGAGRGADPRVPSSPATQVCFBEPQPSTS     180
           |||||||||||||
Db    121  ESLSEKQGMTQESFE------------------------------------------    135

Qy    181  TSDLFPTASA---------STEPSPAGRERSSSCGSGGGGISPKGSGPSVASDEVSSFASA  233
                        |:|      |||||||||||||||||||||||||||||||||||||||
Db    136  --MVPCCVAQAGLKLLGSSDPFTLASQSIVITGSGGGGISPKGSGPSVASDEVSSFASA   192

Qy    234  LPDRKTASFSSLEPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAG  293
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193  LPDRKTASFSSLEPQDQEDLEPEKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAG  252

Qy    294  KPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNH  353
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253  KPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNH  312

Qy    354  ILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDBEGSSEDLLELSD  413
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    313  ILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSD  372

Qy    414  VDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQD  473
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    373  VDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPPTSVSLTTAVQD  432

Qy    474  YVCPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCL  533
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    433  YVCPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCL  492

Qy    534  APFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKQMLTESLVALQRGVFLLSDYR  593
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    493  APFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKQMLTESLVALQRGVFLLSDYR  552

Qy    594  VTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKF  653
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    553  VTGDTVLCYCCGLRSFRELTYQYQQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKF  612

Qy    654  NHICEQTRFKN  664
           |||||||||||
Db    613  NHICEQTRFKN  623
```

```
PR  01-SEP-2000;  2000US-0229345P.
PR  05-SEP-2000;  2000US-0229509P.
PR  05-SEP-2000;  2000US-0229513P.
PR  06-SEP-2000;  2000US-0230437P.
PR  06-SEP-2000;  2000US-0230438P.
PR  08-SEP-2000;  2000US-0231242P.
PR  08-SEP-2000;  2000US-0231243P.
PR  08-SEP-2000;  2000US-0231244P.
PR  08-SEP-2000;  2000US-0231413P.
PR  08-SEP-2000;  2000US-0231414P.
PR  08-SEP-2000;  2000US-0232080P.
PR  08-SEP-2000;  2000US-0232081P.
PR  12-SEP-2000;  2000US-0231968P.
PR  14-SEP-2000;  2000US-0232397P.
PR  14-SEP-2000;  2000US-0232398P.
PR  14-SEP-2000;  2000US-0232399P.
PR  14-SEP-2000;  2000US-0232400P.
PR  14-SEP-2000;  2000US-0232401P.
PR  14-SEP-2000;  2000US-0233063P.
PR  14-SEP-2000;  2000US-0233064P.
PR  14-SEP-2000;  2000US-0233065P.
PR  21-SEP-2000;  2000US-0234223P.
PR  21-SEP-2000;  2000US-0234274P.
PR  25-SEP-2000;  2000US-0234997P.
PR  25-SEP-2000;  2000US-0234998P.
PR  26-SEP-2000;  2000US-0235484P.
PR  27-SEP-2000;  2000US-0235834P.
PR  27-SEP-2000;  2000US-0235836P.
PR  29-SEP-2000;  2000US-0236327P.
PR  29-SEP-2000;  2000US-0236367P.
PR  29-SEP-2000;  2000US-0236368P.
PR  29-SEP-2000;  2000US-0236369P.
PR  29-SEP-2000;  2000US-0236370P.
PR  02-OCT-2000;  2000US-0236802P.
PR  02-OCT-2000;  2000US-0237037P.
PR  02-OCT-2000;  2000US-0237038P.
PR  02-OCT-2000;  2000US-0237039P.
PR  02-OCT-2000;  2000US-0237040P.
PR  13-OCT-2000;  2000US-0239955P.
PR  13-OCT-2000;  2000US-0239937P.
PR  20-OCT-2000;  2000US-0240960P.
PR  20-OCT-2000;  2000US-0241221P.
PR  20-OCT-2000;  2000US-0241785P.
PR  20-OCT-2000;  2000US-0241786P.
PR  20-OCT-2000;  2000US-0241787P.
PR  20-OCT-2000;  2000US-0241808P.
PR  20-OCT-2000;  2000US-0241809P.
PR  20-OCT-2000;  2000US-0241826P.
PR  01-NOV-2000;  2000US-0244617P.
PR  08-NOV-2000;  2000US-0246474P.
PR  08-NOV-2000;  2000US-0246475P.
PR  08-NOV-2000;  2000US-0246476P.
PR  08-NOV-2000;  2000US-0246477P.
PR  08-NOV-2000;  2000US-0246478P.
PR  08-NOV-2000;  2000US-0246523P.
PR  08-NOV-2000;  2000US-0246524P.
PR  08-NOV-2000;  2000US-0246525P.
PR  08-NOV-2000;  2000US-0246526P.
PR  08-NOV-2000;  2000US-0246527P.
PR  08-NOV-2000;  2000US-0246528P.
PR  08-NOV-2000;  2000US-0246532P.
PR  08-NOV-2000;  2000US-0246609P.
PR  08-NOV-2000;  2000US-0246610P.
PR  08-NOV-2000;  2000US-0246611P.
PR  08-NOV-2000;  2000US-0246613P.
PR  17-NOV-2000;  2000US-0249207P.
PR  17-NOV-2000;  2000US-0249208P.
PR  17-NOV-2000;  2000US-0249209P.
PR  17-NOV-2000;  2000US-0249210P.
PR  17-NOV-2000;  2000US-0249211P.
PR  17-NOV-2000;  2000US-0249212P.
PR  17-NOV-2000;  2000US-0249213P.
PR  17-NOV-2000;  2000US-0249214P.
PR  17-NOV-2000;  2000US-0249215P.
PR  17-NOV-2000;  2000US-0249216P.
PR  17-NOV-2000;  2000US-0249217P.
PR  17-NOV-2000;  2000US-0249218P.
PR  17-NOV-2000;  2000US-0249244P.
PR  17-NOV-2000;  2000US-0249245P.
PR  17-NOV-2000;  2000US-0249264P.
PR  17-NOV-2000;  2000US-0249265P.
PR  17-NOV-2000;  2000US-0249297P.
PR  17-NOV-2000;  2000US-0249299P.
PR  17-NOV-2000;  2000US-0249300P.
PR  01-DEC-2000;  2000US-0250160P.
PR  01-DEC-2000;  2000US-0250391P.
PR  05-DEC-2000;  2000US-0251030P.
PR  05-DEC-2000;  2000US-0251988P.
PR  05-DEC-2000;  2000US-0256719P.
PR  06-DEC-2000;  2000US-0251479P.
PR  08-DEC-2000;  2000US-0251856P.
PR  08-DEC-2000;  2000US-0251868P.
PR  08-DEC-2000;  2000US-0251869P.
PR  08-DEC-2000;  2000US-0251989P.
PR  08-DEC-2000;  2000US-0251990P.
PR  11-DEC-2000;  2000US-0254097P.
PR  05-JAN-2001;  2001US-0256678P.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Rosen CA,  Barash SC,  Ruben SM;
XX
DR  WPI; 2001-488783/53.
DR  N-PSDB; AAS25843.
XX
PT  New nucleic acid molecules encoding 461 human secreted proteins for
PT  diagnosing, preventing, treating or ameliorating medical conditions and
PT  used as food additives or preservatives.
XX
PS  Claim 11; SEQ ID NO 809; 980pp; English.
XX
CC  The invention relates to isolated nucleic acid molecules and their
CC  encoded secreted proteins. The nucleic acids and proteins are used to
CC  prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC  rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC  in diagnosing a pathological condition or susceptibility to a
CC  pathological condition. Antibodies to the proteins can also be used in
CC  alleviating symptoms associated with the disorders and in diagnostic
CC  immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC  (ELISA). Disorders which are diagnosed or treated include autoimmune
CC  diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC  neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC  arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC  nervous system disorders e.g. Alzheimer's disease, infections caused by
CC  bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC  and many other disorders listed in the specification. The polypeptides
CC  can also be used to aid wound healing and epithelial cell proliferation,
CC  to prevent skin aging due to sunburn, to maintain organs before
CC  transplantation, for supporting cell culture of primary tissues, to
CC  regenerate tissues and in chemotaxis. The polypeptides can also be used
CC  as a food additive or preservative to increase or decrease storage
CC  capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC  minerals, cofactors and other nutritional components. The present
CC  sequence represents a novel secreted protein of the invention. Note: The
CC  sequence data for this patent did not form part of the printed

Query Match             54.5%;  Score 1954.5;  DB 4;  Length 426;
Best Local Similarity   74.1%;  Pred. No. 3.3e-143;
Matches  389;  Conservative    5;  Mismatches   22;  Indels  109;  Gaps    5;

Qy      1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH  60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH  62

Qy     61 CRIVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY  120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   63 CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHR----- 117

Qy  121 ESLSEKQGMTQESFEANKENVPHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS 180

Db  118 ----------------------------------------------------------- 117

Qy  181 TSDLFPTASASSTEPSPAGRERSSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTA 240
                                  ||||||||||||||||||||||||||
Db  118 ------------------SGGGGISPKGSGPSVASDEVSSFASALPDRKTA 150

Qy  241 SFSSLEPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  151 SFSSLEPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE 210

Qy  301 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  211 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE 270

Qy  361 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSRSFSDEEGSSEDLLELSDVDSESSD 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  271 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSRSFSDEEGSSEDLLELSDVDSESSD 330

Qy  421 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG 480
        ||||||||||||||||||||||||||||||||||||||||          ::
Db  331 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPFTSVSLTT-------VRI 382

Qy  481 SHALC---TCCFQPMPDRRABREQDPRVAPQ-----QCAVCLQPF 517
        :|---|    |    ::        ||
Db  383 TCALCKEATPC-APAASAHARPESGTRTGPRVXXLNKCXGLLQAF 426
```

RESULT 7
ABU54925

ID ABU54925 standard; protein; 426 AA.
XX
AC ABU54925;
XX
DT 18-MAR-2003 (first entry)
XX
DE Human novel polypeptide #12.
XX
KW Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
XX
PN US2002132753-A1.
XX
PD 19-SEP-2002.
XX
PF 17-JAN-2001; 2001US-00764864.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180620P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0209063P.
PR 14-AUG-2000; 2000US-0209648P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0245519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI; 2003-147444/14.
DR N-PSDB; ABX73184.
XX
PT New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.
XX
PS Claim 11; SEQ ID NO 809; 402pp; English.
XX
CC The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC ABU55748 represent human novel polypeptides of the invention
XX
SQ Sequence 426 AA;

Query Match   54.5%; Score 1954.5; DB 6; Length 426;
Best Local Similarity   74.1%; Pred. No. 3.3e-143;
Matches 389; Conservative 5; Mismatches 22; Indels 109; Gaps 5;

```
Qy    1  MEREPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3  MEREPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   62
Qy   61  CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVVRKNEPEHNVAYLY  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63  CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVVRKNEPEHR-----  117
Qy  121  ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPSSPATQVCFEEPQSTS    180
Db  118  --------------------------------------------------------     117
Qy  181  TSDLFPTASASSTEPSPAGRERSSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTA  240
                        ||||||||||||||||||||||||||||||
Db  118  ------------SGGGGISPKGSGPSVASDEVSSFASALPDRKTA  150
Qy  241  SFSSLEPQDQEDLEPVKKQMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE  300
            |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db  151  SFSSLEPQDQEDLEPVKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE   210
Qy  301  TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  211  TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE  270
Qy  361  AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSSSSD  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  271  AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSSSSD  330
Qy  421  ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG  480
            |||||||||||||||||||||||||||||||||               ::
Db  331  ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPFTSVSLTT------VRI   382
Qy  481  SHALC---TCCFQPMPDRRAEREQDPRVAPQ-----QCAVCLQPF  517
            :||    |  |:          |:       :|   |
Db  383  TCALCKEATPC-APASAHARPESGTRTGPRVXXLNKCKGLLQAF  426
```

RESULT 8
AAO08972
ID   AAO08972 standard; protein; 269 AA.
XX
AC   AAO08972;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human polypeptide SEQ ID NO 22864.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US004927.
XX
PR   28-FEB-2000; 2000US-00515126.
PR   18-MAY-2000; 2000US-00577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI88903.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT   and treating e.g. leukemia, inflammation and immune disorders.
XX

PS   Claim 20; SEQ ID NO 22864; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 269 AA;

Query Match   38.0%; Score 1364; DB 4; Length 269;
Best Local Similarity   99.6%; Pred. No. 1.5e-97;
Matches 242; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy  422  SQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGS  481
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27  SQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGS   86
Qy  482  HALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCEINL  541
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   87  HALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCEINL  146
Qy  542  GDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLC  601
            |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db  147  GDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLMALQRGVFLLSDYRVTGDTVLC  206
Qy  602  YCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQTR  661
            |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  207  YCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGNCRTQVKAHHAMKFNHICEQTR   266
Qy  662  FKN  664
            |||
Db  267  FKN  269
```

RESULT 9
AAU16317
ID   AAU16317 standard; protein; 230 AA.
XX
AC   AAU16317;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human novel secreted protein, Seq ID 1270.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
XX

```
PR   24-FEB-2000;   2000US-0184664P.
PR   02-MAR-2000;   2000US-0186350P.
PR   16-MAR-2000;   2000US-0189874P.
PR   17-MAR-2000;   2000US-0190076P.
PR   18-APR-2000;   2000US-0198123P.
PR   19-MAY-2000;   2000US-0205515P.
PR   07-JUN-2000;   2000US-0209467P.
PR   28-JUN-2000;   2000US-0214886P.
PR   30-JUN-2000;   2000US-0215135P.
PR   07-JUL-2000;   2000US-0216647P.
PR   07-JUL-2000;   2000US-0216680P.
PR   11-JUL-2000;   2000US-0217487P.
PR   11-JUL-2000;   2000US-0217496P.
PR   14-JUL-2000;   2000US-0218290P.
PR   26-JUL-2000;   2000US-0220963P.
PR   26-JUL-2000;   2000US-0220964P.
PR   14-AUG-2000;   2000US-0224518P.
PR   14-AUG-2000;   2000US-0224519P.
PR   14-AUG-2000;   2000US-0225213P.
PR   14-AUG-2000;   2000US-0225214P.
PR   14-AUG-2000;   2000US-0225266P.
PR   14-AUG-2000;   2000US-0225267P.
PR   14-AUG-2000;   2000US-0225268P.
PR   14-AUG-2000;   2000US-0225270P.
PR   14-AUG-2000;   2000US-0225447P.
PR   14-AUG-2000;   2000US-0225757P.
PR   14-AUG-2000;   2000US-0225758P.
PR   14-AUG-2000;   2000US-0225759P.
PR   18-AUG-2000;   2000US-0226279P.
PR   22-AUG-2000;   2000US-0226681P.
PR   22-AUG-2000;   2000US-0226868P.
PR   22-AUG-2000;   2000US-0227182P.
PR   23-AUG-2000;   2000US-0227009P.
PR   30-AUG-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0229287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   01-SEP-2000;   2000US-0229345P.
PR   05-SEP-2000;   2000US-0229509P.
PR   05-SEP-2000;   2000US-0229513P.
PR   06-SEP-2000;   2000US-0230437P.
PR   06-SEP-2000;   2000US-0230438P.
PR   08-SEP-2000;   2000US-0231242P.
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.

PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0246474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   N-PSDB; AAS26304.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
     used as food additives or preservatives.
XX
```

PS   Claim 11; SEQ ID NO 1270; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition. Antibodies to the proteins can also be used in
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed

Query Match   23.5%;   Score 843.5; DB 4;   Length 230;
Best Local Similarity   76.3%;   Pred. No. 3.5e-57;
Matches 167;   Conservative   2;   Mismatches 19;   Indels 31;   Gaps   4;

Qy 319 PCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQS 378
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5 PCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQS 64

Qy 379 MDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPEYRRQA 438
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 MDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPEYRRQA 124

Qy 439 AQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQPMP----- 493
         |||||||||||||||||||||||||||| :|
Db 125 AQPPHCPAPEGEPGAPQALGDAPSTSVSLTTV--------RITCALCKEATPCAPAA 173

Qy 494 --DRRAE--REQDPRVA---------PQQCAVCLQPF 517
           ||||| ||||| |||
Db 174 SAHARPESGREQDPRXALXXXPPVLGLHPDRLLRLLAPF 212

RESULT 10
ABU55386
ID   ABU55386 standard; protein; 230 AA.
XX
AC   ABU55386;
XX
DT   18-MAR-2003 (first entry)
XX
DE   Human novel polypeptide #473.
XX
KW   Human; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.

XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA, Ruben SM, Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   N-PSDB; ABX73645.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 11; SEQ ID NO 1270; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.

CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis) allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC ABU55748 represent human novel polypeptides of the invention
XX
SQ Sequence 230 AA;

Query Match   23.5%;   Score 843.5; DB 6; Length 230;
Best Local Similarity 76.3%;   Pred. No. 3.5e-57;
Matches 167; Conservative   2; Mismatches 19; Indels 31; Gaps 4;

Qy 319 PCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQS 378
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5 PCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQS 64

Qy 379 MDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPEYRRQA 438
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 MDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPEYRRQA 124

Qy 439 AQPPHCPAPEGBPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQPMP----- 493
        |||||||||||||||||||||||||||||||| :
Db 125 AQPPHCPAPEGBPGAPQALGDAPSTSVSLTTV----------RITCALCKEATPCAPAA 173

Qy 494 --DRRAB--REQDPRVA----------PQQCAVCLQPF 517
             ||| ||||||| | |||
Db 174 SAHARPESGREQDPRXALXXXPPVLGLHPDRLLRILAPF 212

RESULT 11
ADA54231
ID ADA54231 standard; protein; 128 AA.
XX
AC ADA54231;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human protein, SEQ ID 1799.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-395539/38.
DR N-PSDB; ADA52592.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX

PS Claim 14; SEQ ID NO 1799; 205pp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 128 AA;

Query Match   15.8%;   Score 566; DB 6; Length 128;
Best Local Similarity 99.0%;   Pred. No. 5.9e-36;
Matches 103; Conservative   1; Mismatches   0; Indels 0; Gaps 0;

Qy 561 KNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNI 620
        :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 QNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNI 84

Qy 621 PASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQTRFKN 664
        ||||||||||||||||||||||||||||||||||||||||||||
Db   85 PASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQTRFKN 128

RESULT 12
AAU16156
ID AAU16156 standard; protein; 92 AA.
XX
AC AAU16156;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, Seq ID 1109.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; opthalmological; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001341.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0209063P.
PR 26-JUL-2000; 2000US-0209064P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0254447P.
PR 14-AUG-2000; 2000US-0257757P.
PR 14-AUG-2000; 2000US-0257758P.
PR 14-AUG-2000; 2000US-0257759P.
PR 18-AUG-2000; 2000US-0262279P.
PR 22-AUG-2000; 2000US-0266681P.
PR 22-AUG-2000; 2000US-0266868P.
PR 22-AUG-2000; 2000US-0271182P.
PR 23-AUG-2000; 2000US-0277009P.
PR 30-AUG-2000; 2000US-0289924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0304837P.
PR 06-SEP-2000; 2000US-0304838P.
PR 08-SEP-2000; 2000US-0311242P.
PR 08-SEP-2000; 2000US-0311243P.
PR 08-SEP-2000; 2000US-0311244P.
PR 08-SEP-2000; 2000US-0311413P.
PR 08-SEP-2000; 2000US-0311414P.
PR 08-SEP-2000; 2000US-0320080P.
PR 08-SEP-2000; 2000US-0320081P.
PR 12-SEP-2000; 2000US-0319968P.
PR 14-SEP-2000; 2000US-0323397P.
PR 14-SEP-2000; 2000US-0323398P.
PR 14-SEP-2000; 2000US-0323399P.
PR 14-SEP-2000; 2000US-0324400P.
PR 14-SEP-2000; 2000US-0324401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0238802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.

PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
DR N-PSDB; AAS26143.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
PS Claim 11; SEQ ID NO 1109; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to

CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed

Query Match   13.2%;   Score 474;   DB 4;   Length 92;
Best Local Similarity 98.9%;   Pred. No. 5.3e-29;
Matches   89;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH 62

Qy   61 CRIVVDEKSGQVTLEDTSTSGTVINKLKVV 90
          ||||||||||||||| |||||||||||||
Db   63 CRIVVDEKSGQVTLEDTXTSGTVINKLKVV 92

RESULT 13
ABU55225
ID   ABU55225 standard; protein; 92 AA.
XX
AC   ABU55225;
XX
DT   18-MAR-2003   (first entry)
XX
DE   Human novel polypeptide #312.
XX
KW   Human; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0209963P.
PR   26-JUL-2000; 2000US-0209964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225557P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0289224P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.

PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA, Ruben SM, Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   N-PSDB; ABX73484.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 11; SEQ ID NO 1109; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC   ABU5748 represent human novel polypeptides of the invention
XX
SQ   Sequence 92 AA;

Query Match   13.2%;   Score 474;   DB 6;   Length 92;
Best Local Similarity 98.9%;   Pred. No. 5.3e-29;
Matches   89;   Conservative   0;   Mismatches   0;   Indels   1;   Gaps   0;

Qy   1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH 62

Qy   61 CRIVVDEKSGQVTLEDTSTSGTVINKLKVV 90
          ||||||||||||||| ||||||||||||||
Db   63 CRIVVDEKSGQVTLEDTXTSGTVINKLKVV 92

RESULT 14
AAU16157

```
ID   AAU16157 standard; protein; 99 AA.
XX
AC   AAU16157;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 1110.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.      PR   08-SEP-2000; 2000US-0231413P.
PR   04-FEB-2000; 2000US-0180628P.      PR   08-SEP-2000; 2000US-0231414P.
PR   24-FEB-2000; 2000US-0184664P.      PR   08-SEP-2000; 2000US-0232080P.
PR   02-MAR-2000; 2000US-0186350P.      PR   08-SEP-2000; 2000US-0232081P.
PR   16-MAR-2000; 2000US-0189874P.      PR   12-SEP-2000; 2000US-0231968P.
PR   17-MAR-2000; 2000US-0190076P.      PR   14-SEP-2000; 2000US-0232397P.
PR   18-APR-2000; 2000US-0198123P.      PR   14-SEP-2000; 2000US-0232398P.
PR   19-MAY-2000; 2000US-0205515P.      PR   14-SEP-2000; 2000US-0232399P.
PR   07-JUN-2000; 2000US-0209467P.      PR   14-SEP-2000; 2000US-0232400P.
PR   28-JUN-2000; 2000US-0214886P.      PR   14-SEP-2000; 2000US-0232401P.
PR   30-JUN-2000; 2000US-0215135P.      PR   14-SEP-2000; 2000US-0233063P.
PR   07-JUL-2000; 2000US-0166647P.      PR   14-SEP-2000; 2000US-0233064P.
PR   07-JUL-2000; 2000US-0216880P.      PR   14-SEP-2000; 2000US-0233065P.
PR   11-JUL-2000; 2000US-0217487P.      PR   21-SEP-2000; 2000US-0234223P.
PR   11-JUL-2000; 2000US-0217496P.      PR   21-SEP-2000; 2000US-0234274P.
PR   14-JUL-2000; 2000US-0182990P.      PR   25-SEP-2000; 2000US-0234997P.
PR   26-JUL-2000; 2000US-0220963P.      PR   25-SEP-2000; 2000US-0234998P.
PR   26-JUL-2000; 2000US-0220964P.      PR   26-SEP-2000; 2000US-0235484P.
PR   14-AUG-2000; 2000US-0224518P.      PR   27-SEP-2000; 2000US-0235834P.
PR   14-AUG-2000; 2000US-0224519P.      PR   27-SEP-2000; 2000US-0235836P.
PR   14-AUG-2000; 2000US-0225213P.      PR   29-SEP-2000; 2000US-0236327P.
PR   14-AUG-2000; 2000US-0225214P.      PR   29-SEP-2000; 2000US-0236367P.
PR   14-AUG-2000; 2000US-0225266P.      PR   29-SEP-2000; 2000US-0236368P.
PR   14-AUG-2000; 2000US-0225267P.      PR   29-SEP-2000; 2000US-0236369P.
PR   14-AUG-2000; 2000US-0225268P.      PR   29-SEP-2000; 2000US-0236370P.
PR   14-AUG-2000; 2000US-0225270P.      PR   02-OCT-2000; 2000US-0236802P.
PR   14-AUG-2000; 2000US-0225447P.      PR   02-OCT-2000; 2000US-0237037P.
PR   14-AUG-2000; 2000US-0225757P.      PR   02-OCT-2000; 2000US-0237038P.
PR   14-AUG-2000; 2000US-0225758P.      PR   02-OCT-2000; 2000US-0237039P.
PR   14-AUG-2000; 2000US-0225759P.      PR   02-OCT-2000; 2000US-0237040P.
PR   18-AUG-2000; 2000US-0226279P.      PR   13-OCT-2000; 2000US-0239935P.
PR   22-AUG-2000; 2000US-0226681P.      PR   13-OCT-2000; 2000US-0239937P.
PR   22-AUG-2000; 2000US-0226868P.      PR   20-OCT-2000; 2000US-0240960P.
PR   22-AUG-2000; 2000US-0227182P.      PR   20-OCT-2000; 2000US-0241221P.
PR   23-AUG-2000; 2000US-0227009P.      PR   20-OCT-2000; 2000US-0241785P.
PR   30-AUG-2000; 2000US-0228924P.      PR   20-OCT-2000; 2000US-0241786P.
PR   01-SEP-2000; 2000US-0229287P.      PR   20-OCT-2000; 2000US-0241787P.
PR   01-SEP-2000; 2000US-0229343P.      PR   20-OCT-2000; 2000US-0241808P.
PR   01-SEP-2000; 2000US-0229344P.      PR   20-OCT-2000; 2000US-0241809P.
PR   01-SEP-2000; 2000US-0229345P.      PR   20-OCT-2000; 2000US-0241826P.
PR   05-SEP-2000; 2000US-0229509P.      PR   01-NOV-2000; 2000US-0244617P.
PR   05-SEP-2000; 2000US-0229513P.      PR   08-NOV-2000; 2000US-0246474P.
PR   06-SEP-2000; 2000US-0230437P.      PR   08-NOV-2000; 2000US-0246475P.
PR   06-SEP-2000; 2000US-0230438P.      PR   08-NOV-2000; 2000US-0246476P.
PR   08-SEP-2000; 2000US-0231242P.      PR   08-NOV-2000; 2000US-0246477P.
PR   08-SEP-2000; 2000US-0231243P.      PR   08-NOV-2000; 2000US-0246478P.
PR   08-SEP-2000; 2000US-0231244P.      PR   08-NOV-2000; 2000US-0246523P.
                                        PR   08-NOV-2000; 2000US-0246524P.
                                        PR   08-NOV-2000; 2000US-0246525P.
                                        PR   08-NOV-2000; 2000US-0246526P.
                                        PR   08-NOV-2000; 2000US-0246527P.
                                        PR   08-NOV-2000; 2000US-0246528P.
                                        PR   08-NOV-2000; 2000US-0246532P.
                                        PR   08-NOV-2000; 2000US-0246609P.
                                        PR   08-NOV-2000; 2000US-0246610P.
                                        PR   08-NOV-2000; 2000US-0246611P.
                                        PR   08-NOV-2000; 2000US-0246613P.
                                        PR   17-NOV-2000; 2000US-0249207P.
                                        PR   17-NOV-2000; 2000US-0249208P.
                                        PR   17-NOV-2000; 2000US-0249209P.
                                        PR   17-NOV-2000; 2000US-0249210P.
                                        PR   17-NOV-2000; 2000US-0249211P.
                                        PR   17-NOV-2000; 2000US-0249212P.
                                        PR   17-NOV-2000; 2000US-0249213P.
                                        PR   17-NOV-2000; 2000US-0249214P.
                                        PR   17-NOV-2000; 2000US-0249215P.
                                        PR   17-NOV-2000; 2000US-0249216P.
                                        PR   17-NOV-2000; 2000US-0249217P.
                                        PR   17-NOV-2000; 2000US-0249218P.
                                        PR   17-NOV-2000; 2000US-0249244P.
                                        PR   17-NOV-2000; 2000US-0249245P.
                                        PR   17-NOV-2000; 2000US-0249264P.
                                        PR   17-NOV-2000; 2000US-0249265P.
```

PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.

XX

PS   Claim 11; SEQ ID NO 1110; 980pp; English.

XX

CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed

Query Match   12.1%; Score 432; DB 4; Length 99;
Best Local Similarity   92.5%; Pred. No. 1.1e-25;
Matches   86; Conservative   1; Mismatches   4; Indels   2; Gaps   1;

Qy   42 RRRGCDLSFPSNKLVSGDHCRIVVDEKSGGVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 101
             |||:| | ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9 RDKGPD--FPXNKLVSGDHCRIVVDEKSGGVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 66

Qy   102 VIYLVYRKNEPEHNVAYLYESLSEKQGMTQESF 134
             |||||||||||||||||||||||||||||||||
Db   67 VIYLVYRKNEPEHNVAYLYESLSEKQGMTXESF 99

RESULT 15
AAU16576
ID   AAU16576 standard; protein; 99 AA.
XX
AC   AAU16576;

XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   N-PSDB; AAS26563.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 11; SEQ ID NO 1529; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed

Query Match   12.1%;   Score 432;   DB 4;   Length 99;
Best Local Similarity 92.5%;   Pred. No. 1.1e-25;
Matches 86; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

Qy   42 RRRGCDLSFPSNKLVSGDHCRIVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 101
              |: :| ||||||||||||||||||||||||||||||||||||||||||||
Db   9 RDKGPD--FPXNKLVSGDHCRIVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 66

Qy   102 VIYLVVRKNEPEHNVAYLYESLSEKQGMTQESF 134
              |||||||||||||||||||||||||||| |||
Db   67 VIYLVVRKNEPEHNVAYLYESLSEKQGMTXESF 99

Search completed: May 7, 2004, 14:45:44
Job time : 139.275 secs

OM protein - protein search, using sw model

Run on: May 7, 2004, 14:42:23 ; Search time 44.1732 Seconds
                                     (without alignments)
                                     776.028 Million cell updates/sec

Title: US-10-048-046-2
Perfect score: 3585
Sequence: 1 MERPBEGKQSPPPQPWGRLL.........VKAHHAMKFNHICEQTRFKN 664

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : Issued Patents AA:*
            1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
            2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
            3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
            4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
            5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
            6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	152	4.2	245	2	US-08-897-340-36	Sequence 36, Appl
2	152	4.2	245	2	US-09-252-329-36	Sequence 36, Appl
3	135	3.8	67	1	US-07-945-283-7	Sequence 7, Appl
4	135	3.8	245	2	US-08-897-340-35	Sequence 35, Appl
5	135	3.8	245	2	US-09-252-329-35	Sequence 35, Appl
6	134.5	3.8	826	4	US-09-894-998A-47	Sequence 47, Appl
7	132.5	3.7	2414	5	US-08-227-536-2	Sequence 2, Appli
8	132.5	3.7	2414	5	PCT-US95-04682-2	Sequence 2, Appli
9	131	3.7	1848	3	US-08-296-791-6	Sequence 6, Appli
10	131	3.7	1848	4	US-09-839-996-6	Sequence 6, Appli
11	131	3.7	1848	4	US-10-080-505-6	Sequence 6, Appli
12	131	3.7	1848	5	PCT-US95-10661A-6	Sequence 6, Appli
13	127.5	3.6	989	3	US-08-213-419B-4	Sequence 4, Appli
14	126.5	3.5	933	3	US-08-764-870-14	Sequence 14, Appl
15	126.5	3.5	933	4	US-08-980-115-14	Sequence 14, Appl
16	125.5	3.5	989	4	US-08-213-419B-2	Sequence 2, Appli
17	125	3.5	2482	4	US-09-252-991A-16967	Sequence 16967, A
18	123.5	3.4	1233	4	US-09-688-188B-89	Sequence 89, Appl
19	123.5	3.4	1233	4	US-09-291-417D-89	Sequence 89, Appl
20	123	3.4	395	2	US-08-841-349-9	Sequence 9, Appli
21	123	3.4	395	4	US-09-431-184A-9	Sequence 9, Appli
22	122	3.4	778	6	5198347-4	Patent No. 5198347
23	121.5	3.4	848	1	US-08-045-806-4	Sequence 4, Appli
24	121.5	3.4	848	1	US-08-366-051B-4	Sequence 4, Appli
25	121.5	3.4	1719	2	US-08-459-568-4	Sequence 4, Appli
26	121.5	3.4	1719	2	US-08-399-411-4	Sequence 4, Appli
27	121.5	3.4	1719	3	US-08-516-859A-4	Sequence 4, Appli
28	121.5	3.4	1719	4	US-09-586-472-4	Sequence 4, Appli
29	121.5	3.4	1719	4	US-09-528-706-4	Sequence 4, Appli
30	120.5	3.4	326	2	US-08-786-606-8	Sequence 8, Appli
31	120.5	3.4	797	4	US-09-417-197-141	Sequence 141, App
32	120.5	3.4	808	4	US-08-971-188-12	Sequence 12, Appl
33	120	3.3	984	1	US-08-257-073-3	Sequence 3, Appli
34	120	3.3	984	2	US-08-184-009-120	Sequence 120, App
35	120	3.3	984	2	US-08-458-356-120	Sequence 120, App
36	120	3.3	984	4	US-08-460-736-120	Sequence 120, App
37	120	3.3	984	4	US-09-535-370-120	Sequence 120, App
38	119.5	3.3	381	4	US-09-663-600A-200	Sequence 200, App
39	119	3.3	8991	4	US-08-714-741-32	Sequence 32, Appl
40	117.5	3.3	381	4	US-08-867-057-1	Sequence 1, Appli
41	117.5	3.3	381	2	US-09-128-369-1	Sequence 1, Appli
42	117.5	3.3	605	2	US-08-889-402-2	Sequence 2, Appli
43	117.5	3.3	1190	4	US-09-107-532A-7146	Sequence 7146, Ap
44	117	3.3	223	4	US-09-327-983-5	Sequence 5, Appli
45	116.5	3.2	2441	1	US-08-194-468-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-897-340-36
; Sequence 36, Application US/08897340
; Patent No. 5955306
; GENERAL INFORMATION:
; APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.
; TITLE OF INVENTION: Weight Control Pathway Genes and Uses
; TITLE OF INVENTION: Therefor
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,340
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/715,032
; FILING DATE: 17-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Silveri, Jean M.
; REGISTRATION NUMBER: 39,030
; REFERENCE/DOCKET NUMBER: MNI-005CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 245 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-897-340-36

Query Match   4.2%; Score 152; DB 2; Length 245;
Best Local Similarity   27.4%; Pred. No. 5.3e-05;
Matches   57; Conservative   22; Mismatches   77; Indels   52; Gaps   10;

Qy   286 EDVRAAAGKPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSL-CPTCRC 344

Db 3 EELSAAT--SYTEDDFYCPVCQEVLKTPVRTAACQHVFCRKCFLTAMRESGIHCPLCRG 59

Qy 345 PV---ERICKNHIL-------------NNLVEAYLIQHPDKS----RSBEDVQSMDA 381

Db 60 SVTRRERACPERALDLENIMRFSGSCRCCSKKIKFYRMRHHYKSCKKYQDEYGVSSVIP 119

Qy 382 RNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPE--YRRQ-- 437

Db 120 NFKISQDSV-----RSSNRSETSASDNTETYQEDTSSS--GHPTFKCPLCQESNFTRQRL 172

Qy 438 --------AAQPPHCP---APEGEP 451

Db 173 LDHCNSNHLFQIVPVTCPICVSLPWGDP 200

RESULT 2
US-09-252-329-36
; Sequence 36, Application US/09252329
; Patent No. 6147192
; GENERAL INFORMATION:
; APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.
; TITLE OF INVENTION: Weight Control Pathway Genes and Uses
; TITLE OF INVENTION: Therefor
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: LAHIVE & COCKFIELD, LLP
;   STREET: 28 State Street
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: USA
;   ZIP: 02109
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/252,329
;   FILING DATE:
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/897,340
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Silveri, Jean M.
;   REGISTRATION NUMBER: 39,030
;   REFERENCE/DOCKET NUMBER: MNI-005CP
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617)227-7400
;   TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 36:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 245 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   FRAGMENT TYPE: internal
US-09-252-329-36

Query Match 4.2%; Score 152; DB 3; Length 245;
Best Local Similarity 27.4%; Pred. No. 5.3e-05;
Matches 57; Conservative 22; Mismatches 77; Indels 52; Gaps 10;

Qy 286 EDVRAAAGKPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSL-CPTCRC 344

Db 3 EELSAAT--SYTEDDFYCPVCQEVLKTPVRTAACQHVFCRKCFLTAMRESGIHCPLCRG 59

Qy 345 PV---ERICKNHIL-------------NNLVEAYLIQHPDKS----RSEEDVQSMDA 381

Db 60 SVTRRERACPERALDLENIMRFSGSCRCCSKKIKFYRMRHHYKSCKKYQDEYGVSSVIP 119

Qy 382 RNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPE--YRRQ-- 437

Db 120 NFKISQDSV-----RSSNRSETSASDNTETYQEDTSSS--GHPTFKCPLCQESNFTRQRL 172

Qy 438 --------AAQPPHCP---APEGEP 451

Db 173 LDHCNSNHLFQIVPVTCPICVSLPWGDP 200

RESULT 3
US-07-945-283-7
; Sequence 7, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wesley, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EP0 and LLT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Curtis P. Ribando
;   STREET: 1815 No. 535296th University Street
;   CITY: Peoria
;   STATE: IL
;   COUNTRY: USA
;   ZIP: 61604
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/945,283
;   FILING DATE: 19920911
;   CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
;   NAME: Ribando, Curtis P
;   REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 309-685-4011 ext.513
;   TELEFAX: 309-685-4128
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 67 amino acids
;   TYPE: AMINO ACID
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   FRAGMENT TYPE: internal
;   ORIGINAL SOURCE:
;   ORGANISM: Varicella-zoster virus
US-07-945-283-7

Query Match 3.8%; Score 135; DB 1; Length 67;
Best Local Similarity 36.7%; Pred. No. 0.00022;
Matches 22; Conservative 6; Mismatches 32; Indels 0; Gaps 0;

Qy 303 TCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILANLVEAY 362

Db 5 TCTICMSTVSDLGKTMPCLHDFCFVCIRAWTSTSVQCPLCRCPVQSILHKIVSDTSYKEY 64

RESULT 4
US-08-897-340-35
; Sequence 35, Application US/08897340
; Patent No. 5955306
; GENERAL INFORMATION:
; APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.
; TITLE OF INVENTION: Weight Control Pathway Genes and Uses
; TITLE OF INVENTION: Therefor
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: LAHIVE & COCKFIELD, LLP
;   STREET: 28 State Street
;   CITY: Boston
;   STATE: Massachusetts

```
APPLICATION NUMBER: US/09/252,329
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/897,340
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Silveri, Jean M.
REGISTRATION NUMBER: 39,030
REFERENCE/DOCKET NUMBER: MNI-005CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 245 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-09-252-329-35

Query Match          3.8%;  Score 135;  DB 3;  Length 245;
Best Local Similarity  26.0%;  Pred. No. 0.0015;
Matches  54;  Conservative  23;  Mismatches  79;  Indels  52;  Gaps  10;

Qy   286 EDVRAAAGKPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACY-SGWMERSSLCPTCRC 344
             ||:| ||            : :||::|:|          :  ||||
Db     3 EDLSAAT---SYTEDDFYCPVCQEVLKTPVRTTACQHVFCRKCFLTAMRESGAHCPLCRG 59

Qy   345 PV---ERICKNHILN-----------------NLVEAYLIQHPDKS----RSEEDVQSMDA 381
            |:|   |                       :| |   : :||
Db    60 NVTRRERACPERALDLENIMRKFSGSCRCCAKQIKFYRMRHHYKSCKKYQDEYGVSSIIP 119

Qy   382 RNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPE--YRRQ-- 437
             :|::|                  :          :: |||      :|
Db   120 NFQISQDSVGNSNR-----SETSTSDNTETYQENTSSS--GHPTFKCPLCQESNFTRQRL 172

Qy   438 --------AAQPPHCP---APEGEP 451
                    | |:||
Db   173 LDHCNSNHLFQIVPVTCPICVSLPWGDP 200
```

RESULT 6
US-09-894-998A-47
; Sequence 47, Application US/0894998A
; Patent No. 6537555
; GENERAL INFORMATION:
; APPLICANT: Hosken, Nancy Ann
; APPLICANT: Craig H. Day
; APPLICANT: Davin C. Dillon
; APPLICANT: McGowan, Patrick
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538
; CURRENT APPLICATION NUMBER: US/09/894,998A
; CURRENT FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 826
; TYPE: PRT
; ORGANISM: HSV-2
US-09-894-998A-47

```
Query Match          3.8%;  Score 134.5;  DB 4;  Length 826;
Best Local Similarity  22.0%;  Pred. No. 0.01;
Matches  80;  Conservative  33;  Mismatches  129;  Indels  121;  Gaps  19;

Qy   152 GAGRGADP---RVPPSSPATQVCPEBPQPSTSTSDLFPTASASTEPSPAG------RE 201
Db     6 GTSSRADPGPERPRQTPGTQPA--APHAWGMLNDMQWLASSDSEEETEVGISDDLHRD 63
```

```
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897,340
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/715,032
FILING DATE: 17-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Silveri, Jean M.
REGISTRATION NUMBER: 39,030
REFERENCE/DOCKET NUMBER: MNI-005CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 245 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-897-340-35

Query Match          3.8%;  Score 135;  DB 2;  Length 245;
Best Local Similarity  26.0%;  Pred. No. 0.0015;
Matches  54;  Conservative  23;  Mismatches  79;  Indels  52;  Gaps  10;

Qy   286 EDVRAAAGKPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACY-SGWMERSSLCPTCRC 344
Db     3 EDLSAAT---SYTEDDFYCPVCQEVLKTPVRTTACQHVFCRKCFLTAMRESGAHCPLCRG 59

Qy   345 PV---ERICKNHILN-----------------NLVEAYLIQHPDKS----RSEEDVQSMDA 381
Db    60 NVTRRERACPERALDLENIMRKFSGSCRCCAKQIKFYRMRHHYKSCKKYQDEYGVSSIIP 119

Qy   382 RNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPE--YRRQ-- 437
Db   120 NFQISQDSVGNSNR-----SETSTSDNTETYQENTSSS--GHPTFKCPLCQESNFTRQRL 172

Qy   438 --------AAQPPHCP---APEGEP 451
Db   173 LDHCNSNHLFQIVPVTCPICVSLPWGDP 200
```

RESULT 5
US-09-252-329-35
; Sequence 35, Application US/09252329
; Patent No. 6147192
; GENERAL INFORMATION:
; APPLICANT: Gimeno, Carlos J. and Errada, Patrick, R.
; TITLE OF INVENTION: Weight Control Pathway Genes and Uses
; TITLE OF INVENTION: Therefor
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```
Qy  300 PNMGQQ-PAPQVQQPGLVTPVAQGMSGAHTADPEKRKLIQQQLVLLLHAHKCQRREQAN 358
Db   41 GRRRGCDLSF--PSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPL  97
Db  359 GEVRQCNLPHCRTMKNVLNHMTHCQ--SGKSCQVA--HCASSRQIISHWKNCTRHDCPV 413
Qy   98 ------QTGDVIYLVYRKNEPEHNVAYLYESLSEKQGMTQESFE-ANKENVFHGTKDTSG 150
Db  414 CLPLKNAGD-----KRNQQPILTGAPVGLGNPSSLGVGQQSAPNLSTVSQIDPSSIERAY 468
Qy  151 AGAGRGADPRVPSSPATQVCFEE-PQPSTSTSDLFPTASASTEPSPAGRERSSSCGSG   209
Db  469 AALGLPYQVNQMPTQPQVQAKNQQNQQPGQSPQGMRPMSNMSA---SPMG-----VNG  518
Qy  210 GGGISPKGSGPSVASDEVSSFASALPDR----KTASFSSLEPQD--QEDLEPVKKKMR  261
Db  519 GVGVQT----PSLLSD--SMLHSAINSQNPMSENASVPSLGPMTAAQPSTGIRKQWH   572
Qy  262 GDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPD----KMEETLTCI-ICQDLLH   312
Db  573 EDITQDLRNHL--------VHKLVQAIFPTPDPAALKDRRMENLVAYARKVBGDMYE   621
Qy  313 DCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNN--LVEAYLIQHPDK 369
Db  622 SANNRAEYHLLAEKIYKIQKELEEKRRT----RLQKQNMLPNAAGMVPVSMNPGPNM  675
Qy  370 SRSEBDVQSMD------ARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDIS 422
Db  676 GQPQPGMTSNGPLPDPSMIRGSVPNQMMPRITPQSGLNQFG---------QMSMA    721
Qy  423 QPYVVCRQCP--EYRRQAAQP---------PHCPAPEGE---------P        451
Db  722 QPPIVPRQTPPLQHGQLAQPGALNPPMGYGPRMQQPSNQGQFLPQTQFPSQGMNVTNIP 781
Qy  452 GAPQALGDAPSTSVSLTTAVQDYVCPL------------QGSHALCTCCFQPMPDRRAEREQD 502
Db  782 LAPSS-GQAPVSQAQMSSS----SCPVNSPIMPPGSQGSHIHCPQLPQF-----ALHQNS 831
Qy  503 PRVAPQQ 509
Db  832 PSPVPSR 838
```

```
Qy  202 RSSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTASFSSLEPQDGEDLEPVKKKMR 261
Db   64 STSEAG----------STDTEMFEAGLMDAAT-----------PPARPPAERQG       96
Qy  262 GDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEETLTCIICQDLLHDCVSLQ--P 319
Db   97 SPTPADAQGSC------GGGPVGEEAEAGGGGD------VCAVCTDEIAPPLRCQSFP   143
Qy  320 CMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQSM 379
Db  144 CLHPFCIPCMKTWIPLRNTCPLCNTPV-----------AYLIVGVTASGSFSTIPIV   189
Qy  380 -DARNKITQDMLQPKVRRSFSDEEGSSEDLL----------ELS-----------    412
Db  190 NDPRTRVEAEA----------AVRAGTAVDFIWTGNPRTAPRSLSLGGHTVRALSPTPPWP 240
Qy  413 DVDSESSDISQ--PYY-VCRQCPE-------YRRQAAQPPHC-PAPEGPGAPQ--ALG  458
Db  241 GTDDEDDLADGVDYVPPAPRRAPRGGGGAGATRGTSQPAATRPAP---PGAPRSSSSG  297
Qy  459 DAP 461
Db  298 GAP 300
```

```
Query Match             3.7%;  Score  132.5;  DB 1;  Length 2414;
Best Local Similarity   21.1%;  Pred. No. 0.078;
Matches  128;  Conservative  75;  Mismatches  235;  Indels  169;  Gaps  31;

Qy    4 PEEGKQSPPPQ------PWGRLLRLGAEEGEPH-------VLL-----RKREWTI 40
```

```
;          NAME: Holliday C. Heine, Ph.D.
;          REGISTRATION NUMBER: 34,346
;          REFERENCE/DOCKET NUMBER: DFCI-308Xq999
;          TELECOMMUNICATION INFORMATION:
;          TELEPHONE: (617) 451-0313
;          TELEFAX: (617) 542-2290
;   INFORMATION FOR SEQ ID NO: 2:
;          SEQUENCE CHARACTERISTICS:
;          LENGTH: 2414 amino acids
;          TYPE: amino acid
;          TOPOLOGY: linear
;          MOLECULE TYPE: protein
;PCT-US95-04682-2

Query Match          3.7%;     Score 132.5; DB 5; Length 2414;
Best Local Similarity   21.1%; Pred. No. 0.078;
Matches 128; Conservative 75; Mismatches 235; Indels 169; Gaps  31;

Qy     4 PEEGKQSPPPQ---------PWGRLLRLGAEEGEPH----------VLL---RKREWTI 40
           : :  | |                                        ||||   ||||
Db   300 PNMGQQ-PAPQVQQPGLVTPVAQGMGSGAHTADPEKRKLIQQQLVLLLHAHKCQRREQAN 358

Qy    41 GRRRGCDLSF---PSNKLVSGDHCRIVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPL 97
           ||:|          :  |  ||              : |:  :|  :  | ||
Db   359 GEVRQCNLPHCRTMKQVLNHMTHCQ---SGKSCQVA--HCASSRQIISHWKNCTRHDCPV 413

Qy    98 -----QTGDVIYLVYRKNEPEHNVAYLYRSLSEKQGMTQESFE--ANKENVFHGTKDTSG 150
                           :::    |  | : || :    :|:      :||||||||||
Db   414 CLPLKNAGD----KRNQQPILTGAPVGLGNPSSLGVGQQSAPNLSTVSQIDPSSIERAY 468

Qy   151 AGAGRGADRVPPSSPATQVCFEE-PQPSTSTSDLFPTASASSTEPSPAGRERSSSCGSG 209
           | |  ||| |                : |||   :   ::|:  |      |
Db   469 AALGLPYQVNQMPTQPQVQAKNQQNQQPGQSPQGMRPMSNMSA---SPMG------VNG 518

Qy   210 GGGISPKGSGSPVASDEVSSFASALPDR-----KTASFSSLEPQD--QEDLEPVKKKMR 261
           |  |  || ||| |:  ||| ||:              : |||      ||| |  :
Db   519 GVGVQT----PSLLSD-SMLHSAINSQNPMMSENASVPSLGPMPTAQPSTTGIRKQWH 572

Qy   262 GDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPD-------KMEETLTCI--ICQDLLH 312
           ||             :   :|||||||||||              ||||:|   ||: :
Db   573 EDITQDLRNHL-----------VHKLVQAIFPTPDPAALKDRRMENLVAYARKVEGDMYE 621

Qy   313 DCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNN---LVEAYLIQHPDK 369
           :  | ||  ||      :  |        |    |  | |       |:||| |  |
Db   622 SANNRAEYYHLLAEKIYKIQKELEKRRT-----RLQKQNMLPNAAGMVPVSMNPGPNM 675

Qy   370 SRSEEDVQSMD-------ARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDIS 422
           ||:|| |||         |    : |:                |:|  ||| :|   |
Db   676 GQPQPGMTSNGPLPDPSMIRGSVPNQMPRITPQSGLNQFG----------QMSMA 721

Qy   423 QPYVVCRQCP--EYRRQAAQP---------PHCPAPEGE----------------P 451
           ||:|   |    |  |   |          ||| || |                 |
Db   722 QPPIVPRQTPPLQHGQLAQPGALNPPMCYGPRMQQPSNQGQFLPQTQPPSQGMNVTNIP 781

Qy   452 GAPQALGDAPSTSVSLTTAVQDYVCPL--------QGSHALCTCCFQPMPDRRAEREQD 502
           ||                             |:     |  |:||  |    |
Db   782 LAPSS-GQAPVSQAQMSSS----SCPVNSPIMPPGSQGSHIHCPQLPQP-----ALHQNS 831

Qy   503 PRVAPQQ 509
           :  |
Db   832 PSPVPSR 838

RESULT 9
US-08-296-791-6
; Sequence 6, Application US/08296791
; Patent No. 6245337
;  GENERAL INFORMATION:
;          APPLICANT: St. Geme III, Joseph W.
;          APPLICANT: Falkow, Stanley
;          TITLE OF INVENTION: Haemophilus Adherence and Penetration
;          TITLE OF INVENTION: Protein
;          NUMBER OF SEQUENCES: 9
```

```
;  CORRESPONDENCE ADDRESS:
;          ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;          STREET: 4 Embarcadero Center, Suite 3400
;          CITY: San Francisco
;          STATE: California
;          COUNTRY: United States
;          ZIP: 94111-4187
;  COMPUTER READABLE FORM:
;          MEDIUM TYPE: Floppy disk
;          COMPUTER: IBM PC compatible
;          OPERATING SYSTEM: PC-DOS/MS-DOS
;          SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;          APPLICATION NUMBER: US/08/296,791
;          FILING DATE: 25-AUG-1994
;          CLASSIFICATION: 435
;  ATTORNEY/AGENT INFORMATION:
;          NAME: Trecartin, Richard F.
;          REGISTRATION NUMBER: 31,801
;          REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
;  TELECOMMUNICATION INFORMATION:
;          TELEPHONE: (415) 781-1989
;          TELEFAX:  (415) 398-3249
;          TELEX: 910 277299
;   INFORMATION FOR SEQ ID NO: 6:
;          SEQUENCE CHARACTERISTICS:
;          LENGTH: 1848 amino acids
;          TYPE: amino acid
;          TOPOLOGY: unknown
;US-08-296-791-6

Query Match          3.7%;     Score 131; DB 3; Length 1848;
Best Local Similarity   19.9%; Pred. No. 0.07;
Matches  99; Conservative 67; Mismatches 185; Indels 146; Gaps  22;

Qy    53 NKLVSGDHCRIVDEKSG-----QVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVY 107
           :::|  | |:: |           |:| | |  |  |  | ||||:|  |  |  |:|
Db   929 NKSATGNFTLQVAD-KTGEPNENLTLFDASNA--TRNNLEVT-----LANGSVDRGAW 979

Qy   108 RKNEPEHNVAY-LYESLSEKQGMTQE-----------------SFEANKENVFHGTKDTSGA 151
           :|:|| || |  | :|:||:| |:|                  ||:||||||||||||
Db   980 KYKLRNVNGRYDLYNPEVEKRNQTVDTNITTPNDIQADAPSAQSNNEEIAR--------- 1031

Qy   152 GAGRGADPRVPPSSPATQVCFBPQPSTSTSD-----LFPTASASSTEPSPAGRERSSSC 206
           | || || |  |:|||   |  ||||||| |         | | : | |  |
Db  1032 ----VETPVPPPAPAYBSAIASEQPETRPABTAQPAMEETNTANSTETAPKS-------- 1079

Qy   207 GSGGGGISPKGSGSPVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPVKKKMRGDGDL 266
           |:|| | || || ||| |:  || |                : |||   ||| |  :|
Db  1080 ------------DTATQTENPNSESVPSETTEKVAENPPQBNETV--AKNEQEATEPT 1123

Qy   267 DLNGQLLVA-QPRRNAQTVHEDVRAAAGKPD----------KMEETLTCIICQDLLHDCVSL- 317
           ||      : |  |       | |  |:            ||||| ||| :|  | |  |
Db  1124 PQNGEVAKEDQPTVEANTQTNEATQSEGKTEETQTAETKSEPTESVTVSENQPEKTVSQS 1183

Qy   318 ------------QPCMHTFCAACYSGWMERSSLCP--TCRCP----------VERICKNHILNN 357
                       || ||||||||| |  |: || |  |  |          |  | |  ||
Db  1184 EETQKAPQVTSKEPPKQAEPAPEEVPTD---TN 1230

Qy   358 LVEAYLIQH---------PDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGS 404
           |:|| |         |  : |:|   |:  | |  | | | |:  | | :
Db  1231 AEEAQALGQTQPTTVAAAETTSPNSKPAETQQPSE--KTNAEPVTPVVSENTATQPTE 1287

Qy   405 SEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPG----------- 452
           ||                : :   |:       |:   |    |
Db  1288 TB--ETAKVEKEKTQ-EVPQVASQESPKQEQPAAKPQAQTKPQAEPARENVLTTKNVGE 1343

Qy   453 -APQALGDAPSTSVSLT 468
            ||  | |  || |:
Db  1344 PQPQAQPQTQSTAVPTT 1360

RESULT 10
```

US-09-839-996-6
; Sequence 6, Application US/09839996
; Patent No. 6642371
; GENERAL INFORMATION:
;   APPLICANT: St. Geme III, Joseph W.
;   Falkow, Stanley
;   TITLE OF INVENTION: Haemophilus Adherence and Penetration
;   Protein
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: 4 Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: United States
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/839,996
;   FILING DATE: 20-Apr-2001
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/08/296,791
;   FILING DATE: 25-AUG-1994
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Trecartin, Richard F.
;   REGISTRATION NUMBER: 31,801
;   REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 781-1989
;   TELEFAX: (415) 398-3249
;   TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1848 amino acids
;   TYPE: amino acid
;   TOPOLOGY: unknown
;   SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-839-996-6

Query Match   3.7%; Score 131; DB 4; Length 1848;
Best Local Similarity   19.9%; Pred. No. 0.07;
Matches   99; Conservative   67; Mismatches 185; Indels 146; Gaps   22;

Qy   53 NKLVSGDHCRIVVDEKSG-----QVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVY 107
               :|||| :|:|: |:: |||| :::|:: |:|
Db   929 NKSATGNFTLQVAD-KTGEPNHNELTLFDASNA--TRNNLEVT------LANGSVDRGAW 979

Qy   108 RKNEPEHNVAY-LYESLSEKQGMTQS------------------------SFEANKENVFHGTKDTSGA 151
               : |:|:|||||||:   ::|:
Db   980 KYKLRNVNGRYDLYNPEVEKRNQTVDTTNITTPNDIQADAPSAQSNNEEIAR--------- 1031

Qy   152 GAGRGADPRVPPSSPATQVCFEEPQPSTSTSD----LFPTASASSTEPSPAGRERSSSC 206
               ::|:| ::| ||::|:
Db   1032 -----VETPVPPAPATESAIASEQPETRPAETAQPAMEETNTANSTETAPKS-------- 1079

Qy   207 GSGGGGISPKGSGGPSVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPVKKKMRGDGDL 266
                                ||||| :|
Db   1080 ----------DTATQTENPNSESVPSETTEKVAENPPQENETV--AKNEQEATEPT 1123

Qy   267 DLNGQLLVA-QPRRNAQTVHEDVRAAAGKPD------KMEETLTCIICQDLLHDCVSL- 317
               |:| :|||
Db   1124 PQNGEVAKEDQPTVEANTQTNEATQSEGKTEETQTAETKSEPTESVTVSENQPEKTVSQS 1183

Qy   318 -------QPCMHTFCAACYSGWMERSSLCP--TCRCP------VERICKNHILNN 357
                      :::|: |:|:
Db   1184 TEDKVVVEKEEKAKVET--------EETQKAPQVTSKEPPKQAEPAPEEVPTD---TN 1230

Qy   358 LVEAYLIQH-------------PDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSFSDEEGS 404

US-10-080-505-6
RESULT 11
US-10-080-505-6
; Sequence 6, Application US/10080505
; Patent No. 6676948
; GENERAL INFORMATION:
;   APPLICANT: St. Geme, Joseph W.
;   TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTIENS
;   FILE REFERENCE: A-59941-1/RFT/DCF/DHR
;   CURRENT APPLICATION NUMBER: US/10/080,505
;   CURRENT FILING DATE: 2002-02-22
;   PRIOR APPLICATION NUMBER: US 08/296,791
;   PRIOR FILING DATE: 1994-10-25
;   PRIOR APPLICATION NUMBER: US 09/839,996
;   PRIOR FILING DATE: 2001-04-20
;   NUMBER OF SEQ ID NOS: 58
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
;   LENGTH: 1848
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-10-080-505-6

Query Match   3.7%; Score 131; DB 4; Length 1848;
Best Local Similarity   19.9%; Pred. No. 0.07;
Matches   99; Conservative   67; Mismatches 185; Indels 146; Gaps   22;

Qy   53 NKLVSGDHCRIVVDEKSG-----QVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVY 107
               :|||| :|:|: |:: |||| :::|:: |:|
Db   929 NKSATGNFTLQVAD-KTGEPNHNELTLFDASNA--TRNNLEVT------LANGSVDRGAW 979

Qy   108 RKNEPEHNVAY-LYESLSEKQGMTQS------------------------SFEANKENVFHGTKDTSGA 151
               : |:|:|||||||:   ::|:
Db   980 KYKLRNVNGRYDLYNPEVEKRNQTVDTTNITTPNDIQADAPSAQSNNEEIAR--------- 1031

Qy   152 GAGRGADPRVPPSSPATQVCFEEPQPSTSTSD----LFPTASASSTEPSPAGRERSSSC 206
               ::|:| ::| ||::|:
Db   1032 -----VETPVPPAPATESAIASEQPETRPAETAQPAMEETNTANSTETAPKS-------- 1079

Qy   207 GSGGGGISPKGSGGPSVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPVKKKMRGDGDL 266
                                ||||| :|
Db   1080 ----------DTATQTENPNSESVPSETTEKVAENPPQENETV--AKNEQEATEPT 1123

Qy   267 DLNGQLLVA-QPRRNAQTVHEDVRAAAGKPD------KMEETLTCIICQDLLHDCVSL- 317
               |:| :|||
Db   1124 PQNGEVAKEDQPTVEANTQTNEATQSEGKTEETQTAETKSEPTESVTVSENQPEKTVSQS 1183

Qy   318 -------QPCMHTFCAACYSGWMERSSLCP--TCRCP------VERICKNHILNN 357
                      :::|: |:|:
Db   1184 TEDKVVVEKEEKAKVET--------EETQKAPQVTSKEPPKQAEPAPEEVPTD---TN 1230

Qy   358 LVEAYLIQH-------------PDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSFSDEEGS 404
                                     |: :   ::||:
Db   1231 ABEAQALQQTQPTTVAAAETTSPNSKPAEETQQPSE---KTNAEPVTPVVSENTATQPTE 1287

Qy   405 SEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPG---------- 452
               :|: :|:|: | |: :|:: :|||
Db   1288 TE---ETAKVEKEKTQ-EVPQVASQESPKQEQPAAKPQAQTKPQAEPARENVLTTKNVGE 1343

Qy   453 -APQALGDAPSTSVSLT 468

Db   1344 PQPQAQPQTQSTAVPTT 1360

RESULT 12
PCT-US95-10661A-6
; Sequence 6, Application PC/TUS9510661A
; GENERAL INFORMATION:
;   APPLICANT: Washington University, et al.
;   TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: 4 Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: United States
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/10661A
;   FILING DATE: 16-AUG-1995
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/296,791
;   FILING DATE: 25-AUG-1994
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Trecartin, Richard F.
;   REGISTRATION NUMBER: 31,801
;   REFERENCE/DOCKET NUMBER: FP-59941/RFT
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 781-1989
;   TELEFAX: (415) 398-3249
;   TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1848 amino acids
;   TYPE: amino acid
;   TOPOLOGY: unknown
PCT-US95-10661A-6

Query Match   3.7%;   Score 131; DB 5;   Length 1848;
Best Local Similarity 19.9%;   Pred. No. 0.07;
Matches   99; Conservative   67; Mismatches 185; Indels 146; Gaps   22;

Qy   53 NKLVSGDHCRIVDEKSG-----QVTLEDTSTGTVINKLKVVKKQTCPLQTGDVIYLVY 107
               :||:|| | ||:: | :|| :|:|| ||:|||||:|:::||:|| | |
Db   929 NKSATGNFTLQVAD-KTGEPNHNELITLFDASNA--TRNNLEVT-----LANGSVDRGAW 979

Qy   108 RKNEPEHNVAY-LYESLSEKQGMTQE------------------SFEANKENVFHGTKDTSGA 151
               :|:||||||| : || |::||:   ::| |
Db   980 KYKLRNVNGRYDLYNPEVEKRNQTVDTTNITTPNDIQADAPSAQSNNEEIAR-------- 1031

Qy   152 GAGRGADPRVPPSSPATQVCFEEPQPSTSTSD-----LFPTASASSTEPSPAGRERSSSC 206
               ::||| ||:|:| ::|:| :|:||:|::::||:| :|
Db   1032 -----VETPVPPPAPATESAIASEQPETRPAETAQPAMEETNTANSTETAPKS------- 1079

Qy   207 GSGGGGISPKGSGSPSVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPVKKKMRGDGDL 266
               :: | | |:|:||:| :|:|| | :|:|
Db   1080 ------------DTATQTENPNSESVPSETTEKVAENPPQENETV--AKNEQEATEPT 1123

Qy   267 DLNGQLLVA-QPRRNAQTVHEDVRAAAGKPD------KMEETLTCIICQDLLHDCVSL- 317
               |::: |:   :: | :|:||:||: ::|
Db   1124 PQNGEVAKEDQPTVEANTQTNEATQSEGKTEETCQTAETKSEPTESVTVSENQPEKTVSQS 1183

Qy   318 ----------QPCMCHTFCAACYSGWMERSSLCP--TCRCP-------VERICKNHILNN 357
                         :   :   :
Db   1184 TEDKVVEKEKAKVET----------EETQKAPQVTSKEPPKQAEPAPEEVPTD----TN 1230

Qy   358 LVEAYLIQH------------PDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGS 404

RESULT 13
US-08-213-419B-4
; Sequence 4, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
;   APPLICANT: Inselburg, J. et al.
;   TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
;   TITLE OF INVENTION: AND USES THEREFOR
;   FILE REFERENCE: JII-002CNCP
;   CURRENT APPLICATION NUMBER: US/08/213,419B
;   CURRENT FILING DATE: 1994-03-14
;   PRIOR APPLICATION NUMBER: US 07/870,506
;   PRIOR FILING DATE: 1992-04-17
;   NUMBER OF SEQ ID NOS: 20
;   SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 989
;   TYPE: PRT
;   ORGANISM: Plasmodium falciparum
US-08-213-419B-4

Query Match   3.6%;   Score 127.5; DB 4; Length 989;
Best Local Similarity 20.1%;   Pred. No. 0.055;
Matches   85; Conservative   65; Mismatches 158; Indels 115; Gaps   18;

Qy   49 SFPSNKLVSGDHCRIVDEKSGQVTLEDTSTSG----TVINKLKVVK-----KQTCPLQT 99
               | |||| |||| ||:|| ||||| :|| |::| |||::::| |
Db   63 SEPSNPVSSG-------HSVSTVSVSQTSTSSEKQDTIQVKSALLKDYMGLKVTGPCNE 114

Qy   100 GDVIYLV--YRKNEPEHNVAYLYESLSEKQGMTQESFEANKENV--------FHGTKD 147
               ||:| | ||:|| | || ::||: | :|
Db   115 NFIMFLVPHIYIDVDTEDTNILRLRTTLKETNNAI--SFESNSGSLEKKKYVKLPSNGTVG 172

Qy   148 TSGAGAG--RGADPRVPPSSPATQVCFEEPQPSTSTSDLFPTASASSTEPSPAGRERSSS 205
               : | | ||: | :| | |:| :|:|||:|||||:|| |||:|
Db   173 EQGSSTGTVRGDTEPISDSS---------------SSSSSSSSSSSSSSSSSSSSSSSSSSS 220

Qy   206 CGSGGGGISPKG-SGPSVASDEVSSFASALPDR---------------KT 239
               | | : | ||::| :|||||||:
Db   221 SSSSSESLPANGPDSPTVK------------PPRNLQNICETGKNFKLVYIKENTLLIKW 269

Qy   240 ASFSSLEPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGK--PDK 297
               ::|| | | | : ||:|:| ||:| ::||||| |:: ||
Db   270 KVVGETKDTTEDNKVDVRKYLINEKFPFTSILIHAYKEHNGTNLIIESKNYALGSDIPEK 329

Qy   298 MEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVER-----ICKN 352
               : :|: | | :| |:   :: | :|
Db   330 -------------CDTLASNC--------FLSGNFN--IEK---CFQCALLVEKENKNDVCYK 366

Qy   353 HILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELS 412
               ::: ::|:| : | :| | || :::
Db   367 YLSEDIVSNF---KEIKAETEDDDEDDYTEYKLTESIDNILVKMFKTNENNDKSELIKLE 423

Qy   413 DVD 415

Db   424 EVD 426

RESULT 14
US-08-764-870-14
; Sequence 14, Application US/08764870

Db   1231 AEEAQALIQQTQPTTVAAAETTSPNSKPAEETTQQPSE---KTNAEPVTPVVSENTATQPTE 1287
               :|| :   :   :::| :   : :::
Qy   405 SEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPG---------- 452
               :| :: : :   ::: : :   : :: ||
Db   1288 TB---ETAKVEKETQ-EVPQVASQESPKQEQPAAKPQAQTKPQAEPARENVLTTKNVGE 1343

Qy   453 -APQLGDAPSTSVSLT 468
               :||| :|||:|:|
Db   1344 FQPQAQPQTQSTAVPTT 1360

```
; Patent No. 6235946
; GENERAL INFORMATION:
;   APPLICANT: Scanlan, Thomas S
;   APPLICANT: Baxter, John D
;   APPLICANT: Fletterick, Robert J
;   APPLICANT: Wagner, Richard L
;   APPLICANT: Kushner, Peter J
;   APPLICANT: Apriletti, James W
;   APPLICANT: West, Brian
;   TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand
;   TITLE OF INVENTION: Binding Domains
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Cooley Godward
;     STREET: Five Palo Alto Square, 3000 El Camino Real
;     CITY: Palo Alto
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 94306
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/764,870
;     FILING DATE: 13-DEC-1996
;     CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 60/008,540
;     FILING DATE: 13-DEC-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 60/008,543
;     FILING DATE: 13-DEC-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 60/008,606
;     FILING DATE: 14-DEC-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Nakamura, Jackie N
;     REGISTRATION NUMBER: 35,966
;     REFERENCE/DOCKET NUMBER: UCAL-246/01US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (650)843-5000
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 933 amino acids
;     TYPE: amino acid
;     STRANDEDNESS:
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-764-870-14

Query Match             3.5%;  Score 126.5;  DB 3;  Length 933;
Best Local Similarity  21.3%;  Pred. No. 0.061;
Matches  117;  Conservative  60;  Mismatches 196;  Indels 177;  Gaps  30;

Qy  124  SEKQGWTQESFEANKENVFHGTKDTSGAGAGRGADPR-----------VPPSSPATQV 170
Db   70  SDEKTQDDQQSL-SDVEGAYSRAEATRGAGGSSSSPPEKDSGLLDSVLDTLLAPSGPG--- 125
Qy  171  CFEEPQPS---------TSTSDLF---------PTASASTEPSPAGRERSSSCGSGGG----- 211
Db  126  ---QSQPSPPACEVTSSWCLFGPELPEDDPAAPATQRVLSPLMSRSGCKVGDSSGTAAAH 182
Qy  212  GISPKGSGPS-----VASDEVSSFASALPDRKTASFSSLEPQDQEDLEP--------VKKK 259
Db  183  KVLPRGLSPARQLLLPASE--SPHWSGAPVKPSPQAAAVEVEEEDSSESEESAGPLLKGK 240
Qy  260  MRGDGDLDLNGGLLVAQPRRNAQTV----HEDVRAAAGKPDKMEE-------TLTCIIC 307
Db  241  PRALGGAAAGGAAACPFGAAAGGVALVPKEDSRFSAPRVALVEQDAPMAPGRSPLATTV 300
Qy  308  QDLLHDCVSLQPCMHTFCAA---------CYSGWM--------ERSSLC----------- 339
Db  301  MDPIH-VPILPLNHALLAARTRQLLEDESYDGGAGAASAFAPPRTSPCASSTPVAVGDF 358
Qy  340  PTCRCPVERICKNHILNLVEAYLI----QHPD-KSRSEEDVQSMDARNKIT------- 386
Db  359  PDCAYPPDABPXD------DAYPLYSDFQPPALKIKEBEGARASARSPRSYLVAGANP 411
Qy  387  ---------QDMLQPKVRRSFSDEG--SSEDLLELSDVDSESSDISQPYVVCRQCPEY 434
Db  412  AAFPDFPLGPPPLPPRATPSRPGEAAVTAAPASASVSSASSSGSTL----ECILY 463
Qy  435  RRQAAQP---PHCPAPEGEPGAPQAL---GDAPSTVSLTTA-VQDYVCPLQGSHALCTC 487
Db  464  KAEGAPPQQGPFAPPPCKAPGASGCLLPRDGLPSTSASAAAGAAPALYPALGLNGLPQL 523
Qy  488  CFQP-----------------MPDRRAEREQDPRVA----PQQ-CAVCLQPFCHLYWGC 524
Db  524  GYQAAVLKEGLPQVYPPYLNYLRPD--SEASQSPQYSFESLPQKICLIC------GD 572
Qy  525  TRTGC-YGCL 533
Db  573  EASGCHYGVL 582

RESULT 15
US-08-980-115-14
; Sequence 14, Application US/08980115
; Patent No. 6266622
; GENERAL INFORMATION:
;   APPLICANT: Scanlan, Thomas S.
;   APPLICANT: Baxter, John D.
;   APPLICANT: Fletterick, Robert J.
;   APPLICANT: Wagner, Richard L.
;   APPLICANT: Kushner, Peter J.
;   APPLICANT: Apriletti, James W.
;   APPLICANT: West, Brian L.
;   APPLICANT: Shiau, Andrew K.
;   TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
;   FILE REFERENCE: UCAL-246/02US
;   CURRENT APPLICATION NUMBER: US/08/980,115
;   CURRENT FILING DATE: 1997-11-26
;   EARLIER APPLICATION NUMBER: 08/764,870
;   EARLIER FILING DATE: 1996-12-13
;   EARLIER APPLICATION NUMBER: 60/008,606
;   EARLIER FILING DATE: 1995-12-14
;   EARLIER APPLICATION NUMBER: 60/008,543
;   EARLIER FILING DATE: 1995-12-13
;   EARLIER APPLICATION NUMBER: 60/008,540
;   EARLIER FILING DATE: 1995-12-13
;   NUMBER OF SEQ ID NOS: 17
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 933
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (659)..(918)
;   OTHER INFORMATION: minimal ligand binding domain
US-08-980-115-14

Query Match             3.5%;  Score 126.5;  DB 3;  Length 933;
Best Local Similarity  21.3%;  Pred. No. 0.061;
Matches  117;  Conservative  60;  Mismatches 196;  Indels 177;  Gaps  30;

Qy  124  SEKQGWTQESFEANKENVFHGTKDTSGAGAGRGADPR-----------VPPSSPATQV 170
Db   70  SDEKTQDDQQSL-SDVEGAYSRAEATRGAGGSSSSPPEKDSGLLDSVLDTLLAPSGPG--- 125
Qy  171  CFEEPQPS---------TSTSDLF---------PTASASTEPSPAGRERSSSCGSGGG----- 211
Db  126  ---QSQPSPPACEVTSSWCLFGPELPEDDPAAPATQRVLSPLMSRSGCKVGDSSGTAAAH 182
```

```
Qy  212 GISPKGSGPS------VASDEVSSFASALPDRKTASFSSLEPQDQEDLEP------VKKK 259
            :|:||:           |: :        :|: :::|:             :|:|
Db  183 KVLPRGLSPARQLLLPASE--SPHWSGAPVKPSPQAAAVEVEEEDSSEESAGPLLKGK 240

Qy  260 MRGDGDLDLNGQLLVAQPRRNAQTV----HEDVRAAAGKPDKMEE------TLTCIIC 307
            |||            |:|  |      ||:   :|:  :            :
Db  241 PRALGGAAAGGGAAACPPGAAAGGVALVPKEDSRFSAPRVALVEQDAPMAPGRSPLATTV 300

Qy  308 QDLLHDCVSLQPCMHTFCAA--------CYSGWM-------ERSSLC-------- 339
            :|:||  :|:|     |:           :|:|       ::|
Db  301 MDFIH--VPILPLNHALLAARTRQLLEDESYDGGAGAASAFAPPRTSPCASSTPVAVGDF 358

Qy  340 PTCRCPVERICKNHILNNLVEAYLI----QHPD-KSRSBEDVQSMDARNKIT-------- 386
            |--|: |:      ||:          |: :||: : |:|||
Db  359 PDCAYPPDAEPKD-------DAYPLYSDFQPPALKIKEEEGAEAASARSPRSYLVAGANP 411

Qy  387 ----------QDMLQPKVRRSFSDEEG--SSEDLLELSDVDSESSDISQPYVVCRQCPEY 434
                       |: :       ||:              : :|
Db  412 AAFPDFPLGPPPPLPPRATPSRPGEAAVTAAPASASVSSASSSGSTL------ECILY 463

Qy  435 RRQAAQP---PHCPAPEGEPGAPQAL--GDAPSTSVLTTA-VQDYVCPLQGSHALCTC 487
            : :||    |          |          ||:::   :  |:|||
Db  464 KAEGAPPQQGPFAPPPCKAPGASGCLLPRDGLPSTSASAAAGAPALYPALGLNGLPQL 523

Qy  488 CFQP-------MPDRRAEREQDPRVA----PQQ-CAVCLQPFCHLYWGC 524
            |:|        :||            ||:  :|  :|:|||||:|
Db  524 GYQAAVLKEGLPQVYPPYLNYLRPD--SEASQSPQYSFSLPQKICLIC--------GD 572

Qy  525 TRTGC-YGCL 533
            |:|| ||||
Db  573 EASGCHYGVL 582


Search completed: May 7, 2004, 14:51:21
Job time : 46.1732 secs
```

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   May  7, 2004, 14:49:13 ; Search time 105.174 seconds
                                   (without alignments)
                                   1752.371 Million cell updates/sec

Title:          US-10-048-046-2
Perfect score:  3585
Sequence:     1 MERPEEGKQSPPPQPWGRLL.........VKAHHAMKFNHICEQTRFKN 664

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters:       1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Published_Applications_AA:*
        1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
        2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
        3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
        4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
        5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
        6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
        7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
        8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
        9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
       10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
       11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
       12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
       13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
       14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
       15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
       16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
       17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
       18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	3585	100.0	664	9	US-09-780-525-2	Sequence 2, Appli
2	1954.5	54.5	426	9	US-09-764-864-809	Sequence 809, App
3	1803	50.3	349	15	US-10-108-260A-4516	Sequence 4516, Ap
4	843.5	23.5	230	9	US-09-764-864-1270	Sequence 1270, Ap
5	566	15.8	128	15	US-10-094-749-1799	Sequence 1799, Ap
6	474	13.2	92	9	US-09-764-864-1109	Sequence 1109, Ap
7	432	12.1	99	9	US-09-764-864-1110	Sequence 1110, Ap
8	432	12.1	99	9	US-09-764-864-1529	Sequence 1529, Ap
9	252	7.0	287	12	US-10-424-599-264601	Sequence 264601,
10	146.5	4.1	284	12	US-10-126-103-114	Sequence 114, App
11	146	4.1	303	15	US-10-108-260A-3965	Sequence 3965, Ap
12	142	4.0	648	12	US-10-425-114-63379	Sequence 63379, A
13	139	3.9	775	14	US-10-224-999A-3462	Sequence 3462, Ap
14	139	3.9	825	14	US-10-121-988-161	Sequence 161, App
15	139	3.9	825	14	US-10-200-562-161	Sequence 161, App
16	139	3.9	825	14	US-10-237-551-161	Sequence 161, App
17	139	3.9	825	14	US-10-210-428-1	Sequence 1, Appli
18	139	3.9	4675	15	US-10-093-463-74	Sequence 74, Appl
19	139	3.9	4691	15	US-10-093-463-72	Sequence 72, Appl
20	138	3.8	255	14	US-10-106-698-4383	Sequence 4383, Ap
21	138	3.8	376	12	US-10-424-599-233512	Sequence 233512,
22	138	3.8	420	12	US-10-424-599-160966	Sequence 160966,
23	136.5	3.8	424	12	US-10-425-114-70107	Sequence 70107, A
24	136.5	3.8	521	12	US-10-259-194A-118	Sequence 118, App
25	136	3.8	245	9	US-10-259-194A-118	Sequence 9, Appli
26	135	3.8	285	9	US-09-998-667-9	Sequence 841, App
27	135	3.8	285	9	US-09-764-864-841	Sequence 3125, Ap
28	135	3.8	478	15	US-10-264-049-3125	Sequence 56089, A
29	135	3.8	993	15	US-10-425-114-56089	Sequence 72, Appl
30	135	3.8	353	12	US-1A-259-194A-72	Sequence 187336,
31	134.5	3.8	826	9	US-10-424-599-187336	Sequence 47, Appl
32	134.5	3.8	826	14	US-09-894-998-47	Sequence 47, Appl
33	134.5	3.8	826	14	US-10-121-988-47	Sequence 47, Appl
34	134.5	3.8	826	14	US-10-200-562-47	Sequence 47, Appl
35	134	3.7	331	12	US-10-237-551-47	Sequence 37376, A
36	133.5	3.7	610	15	US-10-425-114-37376	Sequence 2672, Ap
37	133	3.7	346	12	US-10-104-047-2672	Sequence 65516, A
38	132.5	3.7	440	12	US-10-425-114-65516	Sequence 54084, A
39	132.5	3.7	441	12	US-10-425-114-54084	Sequence 203186,
40	132.5	3.7	3507	15	US-10-424-599-203186	Sequence 5784, Ap
41	132	3.7	658	9	US-10-369-493-5784	Sequence 818, App
42	131.5	3.7	191	16	US-09-764-864-818	Sequence 387, App
43	131.5	3.7	563	9	US-10-389-566-387	Sequence 1277, Ap
44	131	3.7	1848	10	US-09-839-996-6	Sequence 6, Appli
45	131	3.7	1848	12	US-10-645-655-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-780-525-2
; Sequence 2, Application US/09780525
; Patent No. US2002000423A1
; GENERAL INFORMATION:
; APPLICANT: Bin-Bing Zhou
; APPLICANT: Yuan Zhu
; APPLICANT: Priya Chaturvedi
; APPLICANT: Mark R. Hurle
; APPLICANT: Xiaotong Li
; TITLE OF INVENTION: FHAR1, A NEW RING FINGER PROTEIN
; FILE REFERENCE: GP-70668-C1
; CURRENT APPLICATION NUMBER: US/09/780,525
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 09/456,876
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 664
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-780-525-2

```
Query Match             100.0%;  Score 3585;  DB 9;  Length 664;
Best Local Similarity   100.0%;  Pred. No. 5e-271;
Matches  664;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH 60

Qy         61 CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY 120

Qy        121 ESLSEKQGMTQESFEANKENVFHGTYKDTSGAGARGADPRVPPSSPATQVCFEBOPOPSTS 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    3 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRGCDLSFPSNKLVSGDH  62
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   61 CRIVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63 CRIVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHR----- 117
Qy  121 ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS 180
Db  118 ------------------------------------------------------------ 117
Qy  181 TSDLFPTASASSTEPSPAGRERSSSCGSGGGISPKGSGPSVASDEVSSFASALPDRKTA 240
        ||||||||||||||||||||||
Db  118 ----------------------SGGGISPKGSGPSVASDEVSSFASALPDRKTA 150
Qy  241 SFSSLEPQDEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  151 SFSSLEPQDEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE 210
Qy  301 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  211 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE 270
Qy  361 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSD 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  271 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSD 330
Qy  421 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG 480
        |||||||||||||||||||||||||||||||||||||||||||||||||        :
Db  331 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPFTSVSLTT-------VRI 382
Qy  481 SHALC---TCCFQPMPDRRAEREQDPRVAPQ-----QCAVCLQPF 517
        ::||     :|| |  :    ::
Db  383 TCALCKEATPC-APAASAHRPESGTRTGPRVXXLNKCXGLLQAF 426
```

```
Db  121 ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS 180
Qy  181 TSDLFPTASASSTEPSPAGRERSSSCGSGGGISPKGSGPSVASDEVSSFASALPDRKTA 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TSDLFPTASASSTEPSPAGRERSSSCGSGGGISPKGSGPSVASDEVSSFASALPDRKTA 240
Qy  241 SFSSLEPQDEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 SFSSLEPQDEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE 300
Qy  301 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE 360
Qy  361 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSD 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSD 420
Qy  421 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG 480
Qy  481 SHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPPCHLYWGCTRTGCYGCLAPFCELN 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 SHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPPCHLYWGCTRTGCYGCLAPFCELN 540
Qy  541 LGDKCLDGVLNNNSYESDILKNYLATRGLTWKQMLTESLVALQRGVFLLSDYRVTGDTVL 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 LGDKCLDGVLNNNSYESDILKNYLATRGLTWKQMLTESLVALQRGVFLLSDYRVTGDTVL 600
Qy  601 CYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQT 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 CYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQT 660
Qy  661 RFKN 664
        ||||
Db  661 RFKN 664
```

RESULT 2
US-09-764-864-809
; Sequence 809, Application US/09764864
; Patent No. US2002013273753A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PTZ23
;   CURRENT APPLICATION NUMBER: US/09/764,864
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 1792
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 809
;   LENGTH: 426
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (414)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (415)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (420)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-809

Query Match   54.5%;   Score 1954.5; DB 9; Length 426;
Best Local Similarity 74.1%;   Pred. No. 5.9e-144;
Matches   389; Conservative   5; Mismatches 22; Indels 109; Gaps 5;

```
Qy    1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRGCDLSFPSNKLVSGDH 60
        ||||
```

RESULT 3
US-10-108-260A-4516
; Sequence 4516, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;   APPLICANT: HELIX RESEARCH INSTITUTE
;   TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
;   FILE REFERENCE: H1-A0106
;   CURRENT APPLICATION NUMBER: US/10/108,260A
;   CURRENT FILING DATE: 2002-03-27
;   NUMBER OF SEQ ID NOS: 5458
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4516
;   LENGTH: 349
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-108-260A-4516

Query Match   50.3%;   Score 1803; DB 15; Length 349;
Best Local Similarity 99.4%;   Pred. No. 3e-132;
Matches   331; Conservative   0; Mismatches   2; Indels 0; Gaps 0;

```
Qy  205 SCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPVKKKMRGDG 264
        :|| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5 SFGSGGGGISPKGSGPSVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPVKKKMRGDG 64
Qy  265 DLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEETLTCIICQDLLHDCVSLQPCMHTF 324
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 DLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEETLTCIICQDLLHDCVSLQPCMHTF 124
Qy  325 CAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQSMDARNK 384
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  125 CAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQSMDARNK 184
Qy  385 ITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHC 444
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  185 ITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHC 244
```

Qy 445 PAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQPMDRRABREQDPR 504
         ||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db 245 PAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQPMDRRVERBQDPR 304

Qy 505 VAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFC 537
         |||||||||||||||||||||||||||||||||
Db 305 VAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFC 337

RESULT 4
US-09-764-864-1270
; Sequence 1270, Application US/09764864
; Patent No. US2002013275SA1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PTZ23
;   CURRENT APPLICATION NUMBER: US/09/764,864
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 1792
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1270
;   LENGTH: 230
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (4)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (189)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (192)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (193)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (194)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (213)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (223)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (229)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (230)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1270

Query Match   23.5%; Score 843.5; DB 9; Length 230;
Best Local Similarity 76.3%; Pred. No. 1.4e-57;
Matches 167; Conservative   2; Mismatches 19; Indels 31; Gaps   4;

Qy 319 PCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEBDVQS 378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5 PCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQS   64

Qy 379 MDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYYVCRCQPEYRRQA 438
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 MDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYYVCRCQPEYRRQA 124

Qy 439 AQPPHCPAPEGBPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQPMP----- 493
         |||||||||||||||||||||||||||||||||||||||||||:
Db 125 AQPPHCPAPEGBPGAPQALGDAPSTSVSLTTV----------RITCALCKEATPCAPAA 173

Qy 494 --DRRAB--REQDPRVA----------PQQCAVCLQPF 517
         | ||||| | ||||| : | ||||||||
Db 174 SAHARPESGREQDPRXALXXXPPVLGLHPDRLLRLLAPF 212

RESULT 5
US-10-094-749-1799
; Sequence 1799, Application US/10094749
; Publication No. US2003021974lA1
; GENERAL INFORMATION:
;   APPLICANT: ISOGAI, TAKAO
;   APPLICANT: SUGIYAMA, TOMOYASU
;   APPLICANT: OTSUKI, TETSUJI
;   APPLICANT: WAKAMATSU, AI
;   APPLICANT: SATO, HIROYUKI
;   APPLICANT: ISHII, SHIZUKO
;   APPLICANT: YAMAMOTO, JUN-ICHI
;   APPLICANT: ISONO, YUUKO
;   APPLICANT: HIO, YURI
;   APPLICANT: OTSUKA, KAORU
;   APPLICANT: NAGAI, KEIICHI
;   APPLICANT: IRIE, RYOTARO
;   APPLICANT: TAMECHIKA, ICHIRO
;   APPLICANT: SEKI, NAOHIKO
;   APPLICANT: YOSHIKAWA, TSUTOMU
;   APPLICANT: OTSUKA, MOTOYUKI
;   APPLICANT: NAGAHARI, KENJI
;   APPLICANT: MASUHO, YASUHIKO
;   TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
;   FILE REFERENCE: 084335/0160
;   CURRENT APPLICATION NUMBER: US/10/094,749
;   CURRENT FILING DATE: 2002-03-12
;   PRIOR APPLICATION NUMBER: 60/350,435
;   PRIOR FILING DATE: 2002-01-24
;   PRIOR APPLICATION NUMBER: JP 2001-328381
;   PRIOR FILING DATE: 2001-09-14
;   NUMBER OF SEQ ID NOS: 3381
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1799
;   LENGTH: 128
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-094-749-1799

Query Match   15.8%; Score 566; DB 15; Length 128;
Best Local Similarity 99.0%; Pred. No. 2.9e-36;
Matches 103; Conservative   1; Mismatches   0; Indels 0; Gaps   0;

Qy 561 KNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNI 620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25 QNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNI   84

Qy 621 PASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQTRPKN 664
         |||||||||||||||||||||||||||||||||||||||||||||
Db   85 PASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQTRFKN 128

RESULT 6
US-09-764-864-1109
; Sequence 1109, Application US/09764864
; Patent No. US2002013275SA1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PTZ23
;   CURRENT APPLICATION NUMBER: US/09/764,864
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - consult PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 1792
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1109
;   LENGTH: 92
;   TYPE: PRT

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1109

Query Match 13.2%; Score 474; DB 9; Length 92;
Best Local Similarity 98.9%; Pred. No. 2.8e-29;
Matches 89; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH 62

Qy 61 CRIVVDEKSGGQVTLEDTSTSGTVINKLKVV 90
        ||||||||||||||||| ||||||||||||
Db 63 CRIVVDEKSGGQVTLEDTXTSGTVINKLKVV 92

RESULT 7
US-09-764-864-1110
; Sequence 1110, Application US/09764864
; Patent No. US2002013275A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1110
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (17)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (96)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1110

Query Match 12.1%; Score 432; DB 9; Length 99;
Best Local Similarity 92.5%; Pred. No. 5.9e-26;
Matches 86; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

Qy 42 RRRGCDLSFPSNKLVSGDHCRIVVDEKSGGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 101
        |:|| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9 RDKGPD--FPXNKLVSGDHCRIVVDEKSGGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 66

Qy 102 VIYLVYRKNEPEHNVAYLYESLSEKQGMTQESF 134
        |||||||||||||||||||||||||| |||
Db 67 VIYLVYRKNEPEHNVAYLYESLSEKQGMTXESF 99

RESULT 8
US-09-764-864-1529
; Sequence 1529, Application US/09764864
; Patent No. US2002013275A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1529

; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (17)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (96)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1529

Query Match 12.1%; Score 432; DB 9; Length 99;
Best Local Similarity 92.5%; Pred. No. 5.9e-26;
Matches 86; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

Qy 42 RRRGCDLSFPSNKLVSGDHCRIVVDEKSGGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 101
        |:|| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9 RDKGPD--FPXNKLVSGDHCRIVVDEKSGGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 66

Qy 102 VIYLVYRKNEPEHNVAYLYESLSEKQGMTQESF 134
        |||||||||||||||||||||||||| |||
Db 67 VIYLVYRKNEPEHNVAYLYESLSEKQGMTXESF 99

RESULT 9
US-10-424-599-264601
; Sequence 264601, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 264601
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_80956C.1.pep
US-10-424-599-264601

Query Match 7.0%; Score 252; DB 12; Length 287;
Best Local Similarity 24.2%; Pred. No. 3.1e-11;
Matches 83; Conservative 39; Mismatches 145; Indels 76; Gaps 10;

Qy 340 PTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFS 399
        || :|| ::| |:
Db   2 PQCRAVVHFAGKNHPLRTIAEDMLRADSSLQRSHDEVALLDT-------YALVRSNLV 52

Qy 400 DEEGSSEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGD 459
                                             |:|
Db 53 IGSGKKQNRKRAYTPLDDQSDG---TYHQCQQC----------------------- 81

Qy 460 APSTSVSLTTAVQDYVCPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCH 519
Db 82 ------VTEVAGFRCKYDTVHLQCQCQACGGWMPSRTGFG-----IPQYCSGCDRSFCG 127

Qy 520 LYW---GCTRTGCYGCLAPFCBLN---LGDKCLDGV---LNNNSYESDILKNYLATRG 568
        |||| || :|:|
Db 128 AYWHALGVTCNGSY----PVCSQDTLRPISDHSISRIPLLAHEKNLHBQNITDSCIRQMG 183

Qy 569 LTWKNMLTESLVALQ------RGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQRQNIPA 622
        ||| :|:|
Db 184 RTLPDVISEWIAKFENREIDRRRVMLNHAEMITARTFVCQDCYHKLVSFLLYWFRLSIPK 243

Qy 623 SELPVAVTSRPDCYWGRNCRTQVKA-HHAMKFNHICEQTRFKN 664

```
Db  244 HLLPPDESAREDCWYGVYACRTQHRSQEHAQKRNHVCRPTRGSN 286
```

RESULT 10
US-10-126-103-114
; Sequence 114, Application US/10126103
; Publication No. US20030224486A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB PATHWA
; FILE REFERENCE: D0108.np
; CURRENT APPLICATION NUMBER: US/10/126,103
; PRIOR APPLICATION NUMBER: US 60/284,962
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/286,645
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: US 60/346,986
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 114
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-126-103-114

Query Match   4.1%; Score 146.5; DB 12; Length 284;
Best Local Similarity   21.5%; Pred. No. 0.0054;
Matches   63; Conservative 38; Mismatches 99; Indels 93; Gaps 13;

```
Qy  164 SSPATQVCFEEP-QPSTSTSDLFPTASAS-STERSP------------------AGR----- 200
            ||:||  :||                                           |||
Db    4 SSSLTRACPPHPRQQGGEWGNKITTKSLGVSHSPSPGTLSETLQSPRNSLREAGRRPAIW 63

Qy  201 -------BRSSSCGSGGGGISPKGSGPS------VASDEVSSFASAL-PDRKTASPSS 244
              :|:   ||  |||                  :|  :|  :  :|
Db   64 TKLRYADADRAALRGEDPGGASSAGSSSQKTDDPERVAGTDCQAFGGGTGSGRLGSSAFKM 123

Qy  245 LEPQDQEDLEPVKKMRGDGDLDLNGQLLVAQPRRN-AQTVH------------ 285
                              ||  :||:               :|: |
Db  124 ASPQ------------------GGQIAIAMRLRNQLQSVYKMDPLRNEVQGRQGYCCG 163

Qy  286 ---EDVRAAGKPDKMEBTLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTC 342
           ::|                ||    :|  :|||  |||:|| |:|  |||
Db  164 RPAEEVRV--KIKDLNEHIVCCLCAGYFVDATTITECIHTFCKSCIVKYLQTSKYCPMC 220

Qy  343 RC------PVERICKNHILNNLVEAYL--IQHPDKSRSBEEDVQSMDARNKITQ 387
           |:        :|  :::  :|: :|      :|:
Db  221 NIKIHETQPILNLKLDRVMQDIVVKLVPGLQDSEKRIREFYQSR-GLDRVTQ 272
```

RESULT 11
US-10-108-260A-3965
; Sequence 3965, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3965
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3965

Query Match   4.1%; Score 146; DB 15; Length 303;
Best Local Similarity   22.4%; Pred. No. 0.0064;

Matches   76; Conservative 33; Mismatches 125; Indels 106; Gaps 20;

```
Qy  302 LTC--IICQDLLHDCVSLQPCMHTFCAACYSGWME---RSSLCPTCRCPVERICKNHILN 356
           :|:  ||  :::| |::| |:|::  : |: : ||     :: ||||
Db   28 ITCKLCLCBQSLDKMTTLQECQCIFCTACLKQYMQLAIRBGCGSPITCP-DMVCLNH--- 83

Qy  357 NLVEAYLIQHPDKSRSBEDVQSMDARNKITQDMLQPKVRRSFSDEGSSEDLLELSDVDS 416
Db   84 ---------------GTLQBABIACLVPVDQFQLYQRLKFERE-----VHL---- 114

Qy  417 BSSDISQPYVVCRQCPEYRRQAAQPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVC 476
Db  115 ------DPYRT-WCPVADCQTV-----CPVASSDPGQP-VLVECPS----------C 148

Qy  477 PLQGSHALCTCCFQPMPDRRABREQDPRVAP----------QQCAVCLQPFCHLY 521
Db  149 HLK----FCSCCKDAWHAEVSCRDSQPIVLFTEHRALFGTDABAPIKQCPVC-RVYIERN 203

Qy  522 WGCTRTGCYGCLAPFCELNLGDKCLDGVLNN---NSYESDILKNYL-ATRGLTW-KNML 575
Db  204 EGCAQMMCKNCKHTFCWY-----CLQNLDNDIPLRHYDKGPCRNKLGHSRASVMWNRTQV 258

Qy  576 TESLVALQRGVFLLSDYRVTGDTVL-----CYCCGLRSFR 610
Db  259 VGILVGL--GIIAL----VTSPLLLLASPCIICCVCKSCR 292
```

RESULT 12
US-10-425-114-63379
; Sequence 63379, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven B
; APPLICANT: Tabaeka, Jack B
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 63379
; LENGTH: 648
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB7319H06_FLI.pep
US-10-425-114-63379

Query Match   4.0%; Score 142; DB 12; Length 648;
Best Local Similarity   19.8%; Pred. No. 0.04;
Matches 102; Conservative 52; Mismatches 169; Indels 192; Gaps 26;

```
Qy  159 PRVPPSSPATQVCFEEPQPSTSTSDLFPTASASSTEPSP------AGRERSSSCG----- 207
            ||:||  :||                                  |||
Db   14 PHPPPSNP--RGCF----ATLSRLLPPRGGRRSDPPDRCSWGSGARRRAABAAQEWNP 66

Qy  208 ------------SGGGGISPKGSGPSVASDEVSSFASALPDRKTASFSSLEPQDQE 251
                        ||  |||                  :|: |
Db   67 RTICTTQTTPPKMTSAAGRLVSP----PATTETQTTTSRTTTPTTPRSS-----SPTDSS 118

Qy  252 DLBPVKKKMRGDGDLD-LNGQLLVAQPR-----RN----AQTVHED------VRAAAGK 294
           :| ||                :||:               :|:
Db  119 ILSEVDIKRRQEDDINRVSTVLSISKPEACVLLRNYNWSVSKVHDEWFADEERVRKVVGL 178

Qy  295 PDK-----MBETLTCIICQDLLHDC----VSLQPCMHTFCAACYSGWMERS-SLCPTCRC 344
           |:|                ||    :|  :|||  |||
Db  179 PEKHIEMPNDREVTCGIC--FESCPLGSMSAAACGHPFCGTCWRGYISTAISDGPGC-- 233

Qy  345 PVERICKNHILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGS 404
```

RESULT 13
US-10-224-999A-3462
; Sequence 3462, Application US/10224999A
; Publication No. US2003017131BA1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Morham, Scott
; APPLICANT: Zavitz, Kenton
; APPLICANT: Hobden, Adrian
; TITLE OF INVENTION: Composition and Method for Treating Viral Infection
; FILE REFERENCE: 5004.01
; CURRENT APPLICATION NUMBER: US/10/224,999A
; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US 60/313,695
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 3484
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3462
; LENGTH: 775
; TYPE: PRT
; ORGANISM: human herpesvirus 1
US-10-224-999A-3462

Query Match   3.9%;   Score 139; DB 14;   Length 775;
Best Local Similarity   22.6%;   Pred. No. 0.088;
Matches 80; Conservative 34; Mismatches 142; Indels 98; Gaps 16;

Qy   156 GADPRVPPSSPATQVCFEEPQPST---STSDLFPTASASSTEPSPAGRERSSSCGSGGG 211
             ::   : : ::::: :::   :   ::: ::::::::: ::   :   :
Db   6 GASTRRPEGRPQ-----REPAPDVWVFPCDRDL-PDSSDSEAETEVGGR------------ 48

Qy   212 GISPKGSGPSVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPVKKMRGDGDLDLNGQ 271
                                                  ::::
Db   49 ------GDADHDD-------------DSASEADSTDTELFETGLLGPQGVD-GGA 84

Qy   272 LLVAQPRRNAQTVHEDVRAAAGKPDK---MEETLTCIICQDLL--HDCVSLQPCMHTFC 325
          :::   ::   :   : :::::   :   ::::::::::::
Db   85 VSGGSPPR-----EEDPGSCGGAPPREDGGSDEGDVCAVCTDEIAPHLRCDTFPCMRFC 139

Qy   326 AACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSR--SEEDVQSMDARN 383
          :::::::::::::::::::::::::::::: ::: :: ::::: : :: :::
Db   140 IPCMKTWMQLRNTCPLCNAKLVYLIVGVTPSGSFSTIPIVNDPQTRMEAEAVRAGTAVD 199

Qy   384 KI----------TQDMLQPKVRRSFSDEEGSSEDLIELSDVDSESSDISQPYV 426
              :::   ::
Db   200 FIWTGNQRFAPRYLTLGGHTVRALSPTHPEPTTDED--------DDDLDDADYVVP-- 247

Qy   427 VCRQCPEY--RRQAAQPP------HCPAPEGEPGAPQA--LGDAPSTSVSLTT 469
          :::::::   ::: :::   :: ::::::::   ::::::::::
Db   248 APRRTPRAPPRRGAAAPPVTGGASHAAPQPAAAARTAPPSAPIGPHGSSVNTTT 301

RESULT 14

US-10-121-988-161
; Sequence 161, Application US/10121988
; Publication No. US20030068327A1
; GENERAL INFORMATION:
; APPLICANT: Hosken, Nancy Ann
; APPLICANT: McGowan, Patrick
; APPLICANT: Sleath, Paul R.
; APPLICANT: Mossman, Sally P.
; APPLICANT: Evans, Lawrence S.
; APPLICANT: Swanson, Ryan M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C1
; CURRENT APPLICATION NUMBER: US/10/121,988
; CURRENT FILING DATE: 2002-04-11
; NUMBER OF SEQ ID NOS: 183
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 161
; LENGTH: 825
; TYPE: PRT
; ORGANISM: HSV2
US-10-121-988-161

Query Match   3.9%;   Score 139; DB 14;   Length 825;
Best Local Similarity   22.7%;   Pred. No. 0.096;
Matches 82; Conservative 33; Mismatches 127; Indels 120; Gaps 19;

Qy   152 GAGRGADP---RVPPSSPATQVCFEEPQPSTSTSDLFPTASASSTEPSPAG------RE 201
              ::   : : ::::: :::   :: :::: ::::::::::::::::   ::
Db   6 GTSSRADPGPERPRQTPGTQPA--APHAWGMLNDMQWLASSDSEETEVGISDDDLHRD 63

Qy   202 RSSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPVKKKMR 261
             :   ::::::
Db   64 STSEAG-----------------STDTEMFEAGLMDAAT----------PPARPPAERQG 96

Qy   262 GDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDRMEEBTLTCIICQDLLHDCVSLQ--P 319
          :::::::::   : :::::::   :: :::::::::   :
Db   97 SPTPADAQGSC-------GGGPVGEEEAEAGGGGD------VCAVCTDEIAPPLRCQSFP 143

Qy   320 CMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQSM 379
          ::::   ::: :::::::::::::::::::::
Db   144 CLHPFCIPCMKTWIPLRNTCPLCNTPV----------AYLIVGVTASGSFSTIPIV 189

Qy   380 -DARNKITQDMI------------------------DEEGSSEDLLEL 411
           ::   : : :
Db   190 NDPRTRVBAEAAVRAGTAVDFIWTGNPRTAPRSLSLGGHTVRALSPTPPWPGTDDEDDDL 249

Qy   412 SDVDSESSDISQPYV--VCRQCPE------YRRQAAQPPHC-PAPEGEPGAPQ--ALGD 459
          ::::   : ::::::   ::::::
Db   250 ADVD--------YVPPAPRRAPRRGGGGAGATRGTSQPAATRPAP---PGAPRSSSSGG 297

Qy   460 AP 461
          ||
Db   298 AP 299

RESULT 15
US-10-200-562-161
; Sequence 161, Application US/10200562
; Publication No. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowen, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C2
; CURRENT APPLICATION NUMBER: US/10/200,562
; CURRENT FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 161
; LENGTH: 825

```
;   TYPE: PRT
;   ORGANISM: HSV2
US-10-200-562-161

Query Match              3.9%;   Score 139;   DB 14;   Length 825;
Best Local Similarity    22.7%;  Pred. No. 0.096;
Matches   82;   Conservative   33;   Mismatches   127;   Indels   127;   Gaps   19;

Qy    152  GAGRGADP---RVPPSSPATQVCFEEPQPSTSTSDLFPTASASSTEPSPAG-------RE 201
                |:|||    :  ||    |          ||  ::  |||:           ::
Db      6  GTSSRADPGPERPPRQTPGTQPA--APHAWGMLNDMQWLASSDSEEETEVGISDDDLHRD 63

Qy    202  RSSSCGSGGGGISPKGSGGPSVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPVKKKMR 261
                :|          :  :   |     :|:    |        :|
Db     64  STSEAG----------STDTEMPEAGLMDAAT----------PPARPPAERQG 96

Qy    262  GDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEETLTCIICQDLLHDCVSLQ--P 319
                ||:     :|          :|:     ::|  |    | :||:|:
Db     97  SPTPADAQGSC------GGGPVGEEEAEAGGGGD------VCAVCTDEIAPPLRCCQSFP 143

Qy    320  CMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQSM 379
                |:||      :           |||              ||||||
Db    144  CLHPFCIPCMKTWIPLRNTCPLCNTPV-----------AYLIVGVTASGSFSTIPIV 189

Qy    380  -DARNKITQDML----------QPKVR-RSFS-----------DEEGSSEDLLEL 411
                |:|||      |:  :      |:  |          |:   :: :|
Db    190  NDPRTRVEABAAVRAGTAVDFIWTGNPRTAPRSLSLGGHTVRALSPTPPWPGTDDEDDDL 249

Qy    412  SDVDSESSDISQPYV--VCRQCPE------YRRQAAQPPHC-PAPEGEPGAPQ--ALGD 459
                :|||         ||   |:||     | ::  :      |||:    ||||:
Db    250  ADVD--------YVPPAPRRAPRRGGGAGATRGTSQPAATRPAP---PGAPRSSSSGG 297

Qy    460  AP 461
                ||
Db    298  AP 299


Search completed: May  7, 2004, 15:06:49
Job time : 107.174 secs
```

```
               GenCore version 5.1.6
       Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   May  7, 2004, 14:41:17 ; Search time 37.1616 Seconds
                                    (without alignments)
                         1718.743 Million cell updates/sec

Title:           US-10-048-046-2
Perfect score:   3585
Sequence:        1 MRRPEEGKQSPPPQPWGRLL.........VKAHHAMKFNHICEQTRFKN 664

Scoring table:   BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched:   283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:   283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  PIR_78:*
            1:  pir1:*
            2:  pir2:*
            3:  pir3:*
            4:  pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1678	46.8	306	2	T46399	hypothetical prote
2	357.5	10.0	473	2	C96516	F16N3.15 [imported
3	168.5	4.7	467	1	W2BE61	gene 61 protein -
4	145	4.0	1603	2	S17983	gene posterior sex
5	144.5	4.0	2322	2	T10542	hypothetical prote
6	142	3.9	532	2	T49467	related to COP1-in
7	141.5	3.9	424	2	G96742	unknown protein F1
8	141.5	3.9	1208	2	T00362	hypothetical prote
9	139	3.9	775	1	EDBEI1	immediate-early pr
10	139	3.9	825	1	EDBEXD	immediate-early pr
11	138	3.8	387	2	T39653	probable DNA repai
12	138	3.8	639	2	T16648	hypothetical prote
13	137.5	3.8	973	2	B89009	protein T27C4.4 [i
14	136.5	3.8	1331	2	T04938	hypothetical prote
15	136.5	3.8	1495	2	A85240	hypothetical prote
16	136.5	3.8	1495	2	T10649	immediate-early pr
17	136	3.8	315	2	D48560	transcription adap
18	132.5	3.7	2414	2	A54277	hypothetical prote
19	132.5	3.7	3507	2	T34513	glioblastoma RING
20	131.5	3.7	551	2	JC7562	Z13 protein - mous
21	131	3.7	794	2	S59069	IgA-specific metal
22	131	3.7	1849	2	C41859	RING-H2 finger pro
23	130.5	3.6	248	2	D96535	hypothetical prote
24	129.5	3.6	453	2	G96695	hypothetical prote
25	129.5	3.6	776	2	T20738	hypothetical prote
26	129.5	3.6	933	1	QRHUP	progesterone recep
27	129.5	3.6	1083	2	C88854	protein F11A10.3 (
28	129	3.6	222	2	JC4296	ring finger protei
29	129	3.6	638	2	JC7753	ring finger B-box
30	129	3.6	1280	2	T00365	hypothetical prote
31	128.5	3.6	1888	2	T14273	zinc finger protei
32	127.5	3.6	245	2	T45652	RNA binding-like p
33	127	3.5	572	2	S21325	probable exo-gluca
34	126.5	3.5	406	2	S59296	probable finger pr
35	126.5	3.5	676	1	EDBB23	immediate-early pr
36	125.5	3.5	568	2	P71614	chromatinic RING f
37	125.5	3.5	989	2	A54505	serine-repeat anti
38	125.5	3.5	2163	2	T15276	hypothetical prote
39	125	3.5	630	2	A49656	estrogen-responsiv
40	124.5	3.5	1576	2	S65774	homeotic protein H
41	124.5	3.5	3869	2	A48205	All-1 protein +GTE
42	124	3.5	676	1	EDBE22	immediate-early pr
43	124	3.5	1712	2	A38261	masking protein pr
44	123.5	3.4	892	2	T09193	ataxin 7 - human
45	123.5	3.4	1233	2	T30989	serine/threonine p

ALIGNMENTS

```
RESULT 1
T46399
hypothetical protein DKFZp434N2420.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46399
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
  submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23031
A;Accession: T46399
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-306 <AAA>
A;Cross-references: EMBL:AL137561
A;Experimental source: adult testis; clone DKFZp434N2420
C;Genetics:
A;Note: DKFZp434N2420.1

Query Match   46.8%;   Score 1678;   DB 2;   Length 306;
Best Local Similarity  100.0%;   Pred. No. 1.9e-100;
Matches 306;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy  359 VEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSES 418
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 VEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSES 60

Qy  419 SDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPL 478
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPL 120

Qy  479 QGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCE 538
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 QGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCE 180

Qy  539 LNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDT 598
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 LNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDT 240

Qy  599 VLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICE 658
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 VLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICE 300

Qy  659 QTRFKN 664
         ||||||
Db  301 QTRFKN 306

RESULT 2
C96516
F16N3.15 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
```

C;Accession: C96516
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
J.; Altafi, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.;
Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.R.; Conway, A.B.; Creasy, T.H.; Dewar, K.;
Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
J.; Koo, H.L.; Mastrangelo, I.A.; Marra, M.A.; Martienssen, R.; McCombie, W.R.; Miguel, T.; Moy, L.; Murphy, B.; Nguyen, M.; O'Shaughnessy, A.L.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
L.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; Van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C96516
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-473 <STO>
A;Cross-references: GB:AE005173; NID:g5668816; PIDN:AAD46042.1; GSPDB:GN00141
C;Genetics:
A;Gene: F16N3.15
A;Map position: 1

Query Match   10.0%;   Score 357.5; DB 2;   Length 473;
Best Local Similarity 23.6%;   Pred. No. 1.8e-15;
Matches   130; Conservative   65; Mismatches 201; Indels 155; Gaps 20;

Qy   177 PSTSTSDLFPTASASSTEPS---------------------------------PAGRERSS 204
              :|:|::| ::|:|:
Db   7 PMTTTAE---TGQSSGSKPSDDDAWAKLVPLDTRFSDIEIRCNDMVICSEIKPSSLEKHE 63

Qy   205 SCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTASFSSLEPQDQDELEPVKKMRGDG 264
              :   :: |:   ::
Db   64 WCR------ITKNLGQSSATIHHKSSDAILVDK-------AVVPK---------DG 97

Qy   265 DLD-LNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEETL----------TCIICQDLLHDC 314
              |: |:| ::| |::|:|| :|| : |:|:|||||||
Db   98 AVDIISGSEIVPGPEEQGYLQYRFTIMPA--PESRTQLLQISIDPEHAKCSICLNIWHDV 155

Qy   315 VSLQPCMHTFCAACYSGWMERSS-------LCPTCRCPVERICKNHILNNLVEAYLIQHPD 368
              |:::|| |:|||| |:|||:||| |||||| |:| ::|||:::| ::||||
Db   156 VTAAPCLHNFCNGCFSEWMRRSEEKHKHVLCPQCRTTVQYVGKNHFLKNIQDEILKVDAA 215

Qy   369 KSRSEEDVQSMDARNKITQDML---QPKVRRSFSDEEGSSEDLLELSDVDSESSSDISQPY 425
              |::||:||:|::|:|:||||| |||||||::|:|:||| |:||:::|:||:::|||
Db   216 LRRPAEDIAVLDSSASIQSNLIIGSKRKRRLNMPAPTHEERDSLRL---------------- 261

Qy   426 VVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALC 485
              ||||:|| |:||| :|:||:||||| |::|:||| |
Db   262 ----QCPQ---------CVANIG----------------GYRCBHHGAHLQC 284

Qy   486 TCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCBLN----L 541
              | :| ::||| |: :| |: |||::: |
Db   285 HLCQGMPFRANLQ------VPLHCKGCDRPFCGAYWS-SENVTQGVSGPVCVRETRPI 337

Qy   542 GDKCLDGV---LNNNSYESDILKNYLATRGLTWKNMLTESL----VALQRGVFLLSDY 592
              | | : ||::::| |::|:||||::|::||:| |::| :|:|
Db   338 SERTITRIPFITHEMNRHEQDITQRCIAHMEKTVPDVVAEWLRLFNNREIDRSRMPLNHA 397

Qy   593 R-VTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQ-VKAHHA 650
              | |:|::|| | ||||||| ||:| |::| |||:|:| |::| |: |: |:||||
Db   398 ETITASTHVCNDCYDKLVGFLLYWFRITLPRNHLPADVAAREDCWIYGYACRTQHHNEDHA 457

Qy   651 MKFNHICEQTR 661
              ||::| |||:|
Db   458 RKRRNVCRPTR 468

RESULT 3
WZBE61
gene 61 protein - human herpesvirus 3
C;Species: human herpesvirus 3, varicella-zoster virus
C;Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 17-Mar-2000
C;Accession: I27215
R;Davison, A.J.; Scott, J.E.
A;Title: The complete DNA sequence of varicella-zoster virus.
J. Gen. Virol. 67, 1759-1816, 1986
A;Reference number: A27345; MUID:86306657; PMID:3018124
A;Accession: I27215
A;Molecule type: DNA
A;Residues: 1-467 <DAV>
A;Cross-references: EMBL:X04370; NID:g59989; PIDN:CAA27944.1; PID:g60050
C;Genetics:
A;Gene: 61
C;Superfamily: varicella-zoster virus gene 61 protein; RING finger homology
C;Keywords: DNA binding; transcription regulation; zinc finger
F;15-63/Domain: RING finger homology <RNG>
F;19-57/Region: zinc finger C3HC4 motif

Query Match   4.7%;   Score 168.5; DB 1;   Length 467;
Best Local Similarity 26.3%;   Pred. No. 0.0024;
Matches   64;   Conservative   20; Mismatches 106; Indels 53; Gaps 10;

Qy   287 DVRAAAGKPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 346
              | :|||:||:||:| || |||:|| ::|||||| |:||| |:|||:||| |||| ||
Db   2 DTILAGGSGTSDASDNTCTICMSTVSDLGKTMPCLHDFCFVCIRAWTSTSVQCPLCRCPV 61

Qy   347 ERICKNHILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSE 406
              |: ::|||::|| :||:| |:   ::|:| ||:   :|||||
Db   62 QSILHKIVSDTSYKEYEV-HP----SDDDGFS----------------EPSFEDSIDILPGDVI 104

Qy   407 DLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVS 466
              :   :: |::||||| ||| :|:||::||
Db   105 DLLPPSPGPSRES-IQQP----------------TSRSSRBPIQSP---NPGPLQSSAREP-TAES 149

Qy   467 LTTAVQDYVCP--LQGSHALCTCC------------FQPMPDRRA----------EREQDPRV 505
              ::||:||||| |:|::| :| |||||||: |||:||||
Db   150 PSDSQQDSIQPPTRDSSPGVTKTCSTASFLRKVFFKDQPAVRSATPVVYGGSIESAQQPRT 209

Qy   506 APQ 508
              | |
Db   210 GGQ 212

RESULT 4
S17983
gene posterior sex combs protein - fruit fly (Drosophila melanogaster)
N;Alternate names: regulatory protein Psc
C;Species: Drosophila melanogaster
C;Date: 22-Nov-1993 #sequence_revision 26-May-1995 #text_change 21-Jul-2000
C;Accession: S17983; S17880
R;Brunk, B.P.; Martin, E.C.; Adler, P.N.
Nature 353, 351-353, 1991
A;Title: Drosophila genes Posterior Sex Combs and Suppressor two of zeste encode protein:
A;Reference number: S17983; MUID:92018190; PMID:1833647
A;Accession: S17983
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1603 <BRU>
A;Cross-references: EMBL:X59275; NID:g8357; PIDN:CAA41965.1; PID:g8358
A;Note: translation of nucleotide sequence is not complete
R;van Lohuizen, M.; Frasch, M.; Wientjens, E.; Berns, A.
Nature 353, 353-355, 1991
A;Title: Sequence similarity between the mammalian bmi-1 proto-oncogene and the Drosophi
A;Reference number: S17880; MUID:92018191; PMID:1922340
A;Accession: S17880
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 231-452,'I',454-591 <LOH>
C;Genetics:
A;Gene: FlyBase:Psc
A;Cross-references: FlyBase:FBgn0005624
C;Superfamily: RING finger homology
C;Keywords: DNA binding; nucleus
F;261-309/Domain: RING finger homology <RNG>

Query Match   4.0%;   Score 145; DB 2;   Length 1603;
Best Local Similarity 20.3%;   Pred. No. 0.35;

```
Matches   84;  Conservative   70;  Mismatches  203;  Indels   56;  Gaps   12;

Qy    67  EKSGQVLEDTSTSGTV----INKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLYES  122
               :::  :| |||  :||  ::::|||
Db    17  KQTAEATPEATATTTMAHTQQKSQLSTLAKTTTTTATNKAAKSVVSNANSSGNNSSKKLA   76

Qy   123  LSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCF--------E  173
              ||:  |
Db    77  LSQSQKTTTTTTTPPTTT----TTTTTAAEATTNAD-KMQKQQQLKQQLFAACSIKVKSE  130

Qy   174  EPQPSTSTSDLFPTASASTEPSPAGRERSSSCG                            233

Db   131  NALATTATPAAL-AAATTTTTATPALATGKAAKTILENGIKKESTPPAVESVEASSSSSS  189

Qy   234  LPDRKTASFSSLEPDQEDLEPVKKMRGDGLDLNGQLLVAQPRRNAQTVHEDVRAAAG    293

Db   190  SSSSSSSSSSWSTTRRATSEDASSNGGASADEEKSEEDPTAAVAASTATTTSDLATTS   249

Qy   294  KPDKMEET------LTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVER  348

Db   250  RPRPVLLTAVNPHIICHLCQGYLINATTIVECLHSFCHSCLINHLRKERFCPRCEMVINN  309

Qy   349  ICKN-------HILNNLVEAYLIQHPD--KSRSEE-DVQSMDARNKITQDMLQPK      393

Db   310  AKPNIKSDTTLQAIVYKLVPGLYEREILMRKRAFYKDRPEEAALATPEQRGDDTEHLI---  366

Qy   394  VRRSFSDEEGSSEDLLELSLSDVDSESSDISQPYVVCRQCPEYRRQAAQPHCPA       446

Db   367  --FSPSDDMSLSLEYAELGELKTD----SEPELVDTLRPRYIQ------CPA        406
```

RESULT 5
T10542
hypothetical protein F3I3.40 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 02-Sep-2000
C;Accession: T10542
R;Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Ba
submitted to the Protein Sequence Database, June 1999
A;Reference number: 217080
A;Accession: T10542
A;Molecule type: DNA
A;Residues: 1-2322 <BEV>
A;Cross-references: EMBL:AL080237; GSPDB:GN00062; ATSP:F3I3.40
A;Experimental source: cultivar Columbia; BAC clone F3I3
C;Genetics:
A;Gene: ATSP:F3I3.40
A;Map position: 4
A;Introns: 320/3; 775/3; 1366/1; 1774/1; 1832/2; 1944/3; 1982/1; 2059/3; 2141/1; 2223/3
C;Superfamily: RING finger homology
F;1914-1961/Domain: RING finger homology <RRN>

```
Query Match     4.0%;  Score 144.5;  DB 2;  Length 2322;
Best Local Similarity   24.2%;  Pred. No. 0.6;
Matches   78;  Conservative   36;  Mismatches  103;  Indels  105;  Gaps   18;

Qy   252  DLEPVKKMRGDGLDLNGQLLVAQ-----PRRNAQTVHEDV---RAAAG-KPDKMEET     301

Db  1502  ELQGIKEKVHG-VDLKLNTRYHVIQVHGSKEMRQEVQKMVNELAREKSALGEKPDEIE--  1558

Qy   302  LTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSS--------LCP-TCRCPVERICK   351

Db  1559  LECPICLSEVDDGYSLEGCSHLFCKACLLEQFEASMRNFDAFPILCSHIDCGAPI-----  1613

Qy   352  NHILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVKRSFSDEEGSSEDLLEL  411

Db  1614  ----VVA-------------------DMRALLSQEKLDELISASLSAFVTSSDGKLRF   1648

Qy   412  SDVDSESSDISQPYVVCRQCPEYRRQAAQPHCPAPEGEPGAPQALGDAPSTVSLTTAV    471

Db  1649  ------------CSTP------DCPSIYRVAG----PQESGEPFICGACHSETCTR-----  1682

Qy   472  QDYVCPLQGSHALCTCCFQPMEPDRRAEREQDPRVA--------PQQCAVCLQPF----  517
```

```
Db  1683  ---CHLE-YHPLITC------ERYKKFKENPDLSLKDWAKGKDVKEPICKSTIEKTDG   1731
             :|:  |||
Qy   518  C-HLYWGCTRTGCYGCLAPFCE  538
              :|:  |||:
Db  1732  CNHLQCRCGKHICWTCLDVFTQ  1753
```

RESULT 6
T49467
related to COPi-interacting protein CIP8 [imported] - Neurospora crassa
N;Alternate names: protein B14D6.190
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Sep-2000
C;Accession: T49467
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25022
A;Accession: T49467
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-532 <SCH>
A;Cross-references: EMBL:AL356173; GSPDB:GN00116; NCSP:B14D6.190
A;Experimental source: BAC clone B14D6; strain OR74A
C;Genetics:
A;Gene: NCSP:B14D6.190
A;Map position: 6
C;Superfamily: RING finger homology
F;418-468/Domain: RING finger homology <RRN>

```
Query Match     4.0%;  Score 142;  DB 2;  Length 532;
Best Local Similarity   20.0%;  Pred. No. 0.14;
Matches   86;  Conservative   60;  Mismatches  163;  Indels  120;  Gaps   17;

Qy    73  TLEDTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEH-NVAYLYESL---------  123

Db   153  TSDETNTRGT-------QEQTSNASNTEAPRSGSRSTNTQHFDIRFIFPPMTFTTVIS   204

Qy   124  -----------SEKQGMTQESFEANKE-----NVFHGTKDTS------GAGAGRGADPRV-  161

Db   205  EPLTPRSQQQPAGSQPNGQAQSTEQSNEQPNNQQSNASNTSTNVFPTPNGSTASGPSPITF  264

Qy   162  -------PP----SSPATQVCFEEFQPSTSTSDLFPTASASTEPSPAGRERSSSCG--    207

Db   265  FGLHFFTPRPPHAHTAPASNAA------SNTTQTEGTAQQQSQSNDEQAQASQQQPQCGPQ  320

Qy   208  -SGGGGISPKGSGPSVASDEVSSFASALPDRKTASPSSLEPQDQBDLEPVKKKMRGDGDL  266

Db   321  PATQGQQSPFGSGP-VPAGLMNAILSAM-------FSPLSPLVLGNLNFVNDGVYSQBAF  372

Qy   267  D-LNGQLLVAQPRRNAQTVHEDVRAAAGK-------PDKMEETL-------TCIICQD--LLH  312

Db   373  DRIITQLREQHAAQNPGGAPPASQAAIEKLRVKDIDEQMLQGCQDNKAKCVICVDEMTLG  432

Qy   313  DCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRS  372

Db   433  DKATLLPCNHFFHGECVTPWLKVHNTCPVCRRSVE--------DEQQSAFVAAAIP     467

Qy   373  EEDVQSMDARNKITQDMLQPKVRRSFSDEGGSSEDLLELSDVDSESSDISQPYVVCRQCP  432

Db   468  ------VEEAPESKKRKVAEHEPTGRDGHGIDEAEDE--DEQQSAFVAAAIP        512

Qy   433  EYRRQAAQP  441

Db   513  ERQRRPVGP  521
```

RESULT 7
G96742
unknown protein F17M19.13 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: G96742

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:1130712
A;Accession: G96742
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-424 <STO>
A;Cross-references: GB:AE005173; NID:g6978922; PIDN:AAF34314.1; GSPDB:GN00141
C;Genetics:
A;Gene: F17M19.13
A;Map position: 1

Query Match   3.9%;   Score 141.5; DB 2;   Length 424;
Best Local Similarity 26.3%;   Pred. No. 0.11;
Matches   67; Conservative   27; Mismatches 108; Indels   53; Gaps   12;

Qy   268 LNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEETLTCIICQD--LLHDCVSLQPCMHTFC 325
             | :::| | |||||||||| :| || ||| ||| |:|| |||
Db   182 MSRRLVKAMPSLIFSSFHEDNTTA---------FTCAICLEDYTVGDKLRLLPCCHKFH 231

Qy   326 AACYSGWMER-SSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQSMDARNK 384
             |||||||| |:||||||| | |: |||| ||| ||::
Db   232 AACVDSWLTSWRTFCP--------VCKRDARTSTGE------PPASESTPLLSS--AASS 275

Qy   385 ITQDMLQPKVRR--------------SFSDEEGSSEDLLELSDVDSESSDISQPYVVC 428
             | | :||| | | :: || :||| || :::| | |||
Db   276 FTSSSLHSSVRSSALLIGPSLGSLPTISIFSSPAYASSSYIRQ--SFQSSSNRRSPPISVS 333

Qy   429 RQCPEYRRQAAQPPHCPAPE--GEPGAPQALGDAPSTSVSLTTAVQDY-VCPLQGSHAL 484
             ::||::| ||:| :|| || |: || ::|| | |:|
Db   334 RSSVDLRQQAASPSPSPSPSQRSYISHMASPQSLG-YPTISPFNTRYMSPYRPSPSNASPAM 392

Qy   485 CTCCFQPM-PDRRAE 498
             :| | | | ||:
Db   393 AGSSNYPLNPLRYSE 407

RESULT 8
T00362
hypothetical protein KIAA0675 - human
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 02-Sep-2000
C;Accession: T00362
R;Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A;Reference number: Z14142; MUID:98403880; PMID:9734811
A;Accession: T00362
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1208 <ISH>
A;Cross-references: EMBL:AB014575; NID:g3327163; PIDN:g3327163; PIDN:BAA31650.1; PID:g3327164
A;Experimental source: brain; clone HK02566
C;Genetics:
A;Note: KIAA0675
C;Superfamily: RING finger homology
F;1144-1193/Domain: RING finger homology <RRN>

Query Match   3.9%;   Score 141.5; DB 2;   Length 1208;
Best Local Similarity 26.1%;   Pred. No. 0.42;
Matches   54; Conservative   28; Mismatches   92; Indels   33; Gaps   8;

Qy   159 PRVPPSSPATQVCFEEPQPSTSTSDLFPTASASSTEPSPAGRERSSSCGSGGGGISPKGS 218
             | |:   ::| || :   :
Db   999 PRAP--LMTGIAWALPAP---VGDAVPPSAGLRSDPSIMNWERITD-------RLKTA 1044

Qy   219 GPSVASDEVSSFASALPDRKTASFSSLEPQD-----QEDLEPVKKKMRGDGDLDLN---- 269
             ||||:|||||||||||||||||||||||||| ||||||||||||||| ::|
Db 1045 FPQQTRKELTDFLRKLKDAYGKSLSELTFDEIVCKISQFIDPKKSQSOGKSVSNVNCVSP 1104

Qy   270 ----GQLLVAQPRRNAQTVHEDVRAA----AGKPD-KMEETLTCIIC-QDLLHDCVSLQP 319
                 ||||||||||||||||||||| : |::||| :: ||||||||||
Db 1105 SHSPSQPDAAQPPKPAWRPLTSQGPATWEGASNPDEEEEEEPVICHENLSPENLSVLP 1164

Qy   320 CMHTFCAACYSGWMERSSLCPTCRCPV 346
             ||| ||||||||||||||||||||||
Db 1165 CAHKFHAQCIRPWLMQQGTCPTCRLHV 1191

RESULT 9
EDBB11
immediate-early protein IE110 - human herpesvirus 1 (strain 17)
C;Species: human herpesvirus 1
C;Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 17-Mar-2000
C;Accession: A29152
R;Perry, L.J.; Rixon, F.J.; Everett, R.D.; Frame, M.C.; McGeoch, D.J.
J. Gen. Virol. 67, 2365-2380, 1986
A;Title: Characterization of the IE110 gene of herpes simplex virus type 1.
A;Reference number: A29152; MUID:87059760; PMID:3023529
A;Accession: A29152
A;Molecule type: DNA
A;Residues: 1-775 <PER>
A;Cross-references: GB:X04614; NID:g59832; PIDN:CAA28285.1; PID:g59833
C;Genetics:
A;Introns: 19/3; 242/1
C;Superfamily: herpesvirus immediate-early protein IE110; RING finger homology
C;Keywords: DNA binding; early protein; transcription regulation; zinc finger
F;112-162/Domain: RING finger homology <RNG>
F;116-156/Region: zinc finger C3HC4 motif

Query Match   3.9%;   Score 139; DB 1;   Length 775;
Best Local Similarity 22.6%;   Pred. No. 0.35;
Matches   80; Conservative   34; Mismatches 142; Indels   98; Gaps   16;

Qy   156 GADPRVPPSSPATQVCFEBPQPST----STSDLFPTASASSTEPSPAGRERSSSCGSGGG 211
             | :| || | ||| | |:| :|| ||||
Db   6 GASTRRPEGRPQ-----REPAPDVWVFPCDRDL-PDSSDSEAETTEVGGR---------- 48

Qy   212 GISPKGSGSPSVASDEVSSFASALPDRKTASFSSLEPQDOEDLEPVKKKMRGDGDLDLNGQ 271
             | :: |::   :: |||| | |:||: | | |
Db   49 G--------GDADHHDD--------------------DSASEADSTDTELFETGLLGPQGVD-GGA 84

Qy   272 LLVAQPRRNAQTVHEDVRAAAGKPDK----MEETLTCIICQDLL--HDCVSLQPCMHTFC 325
             ||||||||||||||||||||||| |:|||: | ||||||| ||||| |
Db   85 VSGGSPPR-----EEDPGSCGGAPPREDGGSDEGDVCAVCTDEIAPHLRCDTFPCMHRFC 139

Qy   326 AACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSR--SEEDVQSMDARN 383
             |||: | ||| ||||||||||||||::|:||| |:| |::|:   ::||||:|:
Db   140 IPCMKTWMQLRNTCPLCNAKLVYLIVGVTPSGSFSTIPIVNDPQTRMEAEEAVRAGTAVD 199

Qy   384 KI-----------------------TQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYV 426
             || | : |: |:
Db   200 FIWTGNQRFAPRYLLLGGHTVRALSPTHEPTTDED-------DDDLDDADYVPP-- 247

Qy   427 VCRQCPEY--RRQAAQPP-------HCPAPEGEPGAPQA--LGDAPSTSVSLTT 469
             | : | | || | | ||: | :|| |
Db   248 APRRTRAPPRGAAAPPVTGGASHAAPQPAAARTAPPSAPIGPHGSSNTNTTT 301

RESULT 10
EDBEXD
immediate-early protein RL2 - human herpesvirus 2 (strain HG52)
N;Alternate names: RL2 protein
C;Species: human herpesvirus 2
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
C;Accession: JQ1501
R;McGeoch, D.J.; Cunningham, C.; McIntyre, G.; Dolan, A.

```
J. Gen. Virol. 72, 3057-3075, 1991
A;Title: Comparative sequence analysis of the long repeat regions and adjoining parts of
A;Reference number: JQ1494; MUID:92113549; PMID:1662697
A;Accession: JQ1501
A;Molecule type: DNA
A;Residues: 1-825 <MCG>
A;Cross-references: GB:D10471; DDBJ:D01128; NID:g221784; PIDN:BAA23427.1; PID:g2626942
C;Genetics:
A;Gene: RL2
A;Introns: 25/3; 252/1
C;Superfamily: herpesvirus immediate-early protein IE110; RING finger homology
C;Keywords: DNA binding; immediate-early protein; tandem repeat; transcription regulatio
F;122-172/Domain: RING finger homology <RNG>
F;126-166/Region: zinc finger C3HC4 motif
F;589-621/Region: 5-residue repeats (A-S-S-S)

Query Match          3.9%;  Score 139;  DB 1;  Length 825;
Best Local Similarity  22.7%; Pred. No. 0.38;
Matches  82;  Conservative  33;  Mismatches 127;  Indels 120;  Gaps  19;

Qy 152 GAGRGADP--RVPPSSPATQVCFEEPQPSTSTSDLFPTASASSTEPSPAG-------RE 201
Db   6 GTSSRADPGPERPPRQTPGTQPA--APHAWGMLNDMQWLASSDSEEETEVGISDDDLHRD 63

Qy 202 RSSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPVKKQMR 261
Db  64 STSEAG------------STDTEMFEAGLMDAAT----------PPARPPAERQG 96

Qy 262 GDGDLDLNGQLLVAQPRRNACTVHEDVRAAAGKPDKMEETLTCIICQDLLHDCVSLQ-P 319
Db  97 SPTPADAQGSC------GGGPVGEEEAEAGGGGD------VCAVCTDEIAPPLRCQSPP 143

Qy 320 CMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQSM 379
Db 144 CLHPFCIPCMKTWIPLRNTCPLCNTPV------------AYLIVGVTASGSFSTIPIV 189

Qy 380 -DARNKITQDML-----------------QPKVR-RSFS------------DEEGSSEDLLEL 411
Db 190 NDPRTRVEAEAAVRAGTAVDFIWTGNPRTAPRSLSLGGHTVRALSPTPWPGTDDBDDDL 249

Qy 412 SDVDSESSDISQPYV--VCRQCPE------YRRQAAQPPHC-PAPEGERPGAPQ--ALGD 459
Db 250 ADVD-----YVPPAPRRAPRRGGGGAGATRGTSQPAATRPAP---PGAPRSSSSGG 297

Qy 460 AP 461
Db 298 AP 299


RESULT 11
T39653
probable DNA repair and recombination protein - fission yeast (Schizosaccharomyces pombe
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C;Accession: T39653
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z21868
A;Accession: T39653
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-387 <LYN>
A;Cross-references: EMBL:AL031856; PIDN:CAA21300.1; GSPDB:GN00067; SPDB:SPBC1734.06
A;Experimental source: strain 972h-; cosmid c1734
C;Genetics:
A;Gene: SPDB:SPBC1734.06
A;Map position: 2
C;Superfamily: RING finger homology
F;25-72/Domain: RING finger homology <RRN>

Query Match          3.8%;  Score 138;  DB 2;  Length 387;
Best Local Similarity  25.2%; Pred. No. 0.17;
Matches  38;  Conservative  28;  Mismatches  71;  Indels  14;  Gaps  6;
```

```
Qy 298 MEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVE--RICKNHIL 355
Db  23 LDSSLRCLICHEYFR-APLITSCSHTFCSFCIRDYLREHPMCPACRAPEQESRLRKNTIL 81

Qy 356 NNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVD 415
Db  82 EEILESFKVIRPTL-----FEFLKVEN-VPKPVLQAPETVIAQDSASGDEEWED--DLA 132

Qy 416 SESSDISQPYVVCRQCPEYRRQAAQPPHCPA 446
Db 133 SNSSPASIAKKTSRDSKKKRE--DLVHCPA 161


RESULT 12
T16648
hypothetical protein R02E12.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Sep-2000
C;Accession: T16648
R.Leimbach, D.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid R02E12.
A;Reference number: Z18554
A;Accession: T16648
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-639 <LEI>
A;Cross-references: EMBL:U53337; NID:g1255833; PID:g1255835; PIDN:AAA96184.1; GSPDB:GN00(
A;Experimental source: strain Bristol N2; clone R02E12
C;Genetics:
A;Gene: CESP:R02E12.4
A;Map position: X
A;Introns: 63/3; 90/1; 136/2; 177/1; 220/2; 242/3; 294/2; 427/2; 464/3; 505/3; 580/1
C;Superfamily: RING finger homology
F;22-70/Domain: RING finger homology <RRN>

Query Match          3.8%;  Score 138;  DB 2;  Length 639;
Best Local Similarity  28.8%; Pred. No. 0.32;
Matches  44;  Conservative  21;  Mismatches  66;  Indels  22;  Gaps  7;

Qy 283 TVHEDVRAAAGKPD--KME-ETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLC 339
Db   2 TVEEQVVEMLSNLDFESLEFEDVQCHICFQVNHEPVTFLTCKHSICAGCAGRWLSSCSVC 61

Qy 340 PTCRCPVERICKNHILNNLVEAYLIQHPDKSRSEEDVQ----------SMDARNKITQ- 387
Db  62 PMCRTRVREIHLNDDLKKKSLEFLKLHPEQ-ELPEDIKWEQCCHETVFWSLQQKRANES 120

Qy 388 DML----QPKVRRSFSDEEGSSEDLLELSDVDS 416
Db 121 DAVFYGSQPVVPEHLSAEB--SNDFIETKRLEN 151


RESULT 13
B89009
protein T27C4.4 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: B89009
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans; a platform for investigating biology
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans] and www.sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: B89009
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-973 <STO>
A;Cross-references: GB:chr_V; PIDN:AAC17700.1; PID:g3165588; GSPDB:GN00023; CESP:T27C4.4
C;Genetics:
A;Gene: T27C4.4
```

A;Map position: 5

```
Query Match             3.8%;    Score 137.5;  DB 2;  Length 973;
Best Local Similarity  22.4%;    Pred. No. 0.58;
Matches   95;  Conservative   55;  Mismatches  190;  Indels  85;  Gaps  18;

Qy     72 VTLEDTSTSGT--VINKL-KVVKKQTCPLQTGVIYLVYRKNEPEHNVAYLYESLSEKQG 128
          : :  ||: |   :|  |  ||
Db     21 VYFDDTSATDAPYQIRKIEELVKTEKGGVDARCVVYL-RRRDIPQH----LLKIADQAQR 75

Qy    129 MTQESFEANKENVFHGTKDTSGAGARGADPRVPPSSP-ATQVCFE-EPQPSTSTSDLF 185
          :  |: :||      :|      :|  |    |:  |   |
Db     76 RFDNYYEVDKKKPENFT--TKGFIVVNGASEAAPEAAPEATEASAETSETKVKEDTDDVA 133

Qy    186 PTASASSTEPSPAGRERSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTASFSSL 245
          |        | |  | ||              ::   ||| |
Db    134 EKMEDEECPAPVGQGRTSST--------EEPS-ASDAPATAKDKKEEKEAKENAK 181

Qy    246 EPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQ------TVHEDVRAAGKPDKM- 298
          |   |:|   :    | ||:|||        :             ||||||| |||||
Db    182 NLKETABEAEATKLIDWGDGGLPLGVDKLTPDQRLKLRQHEIFMTRQSEILPAAAIRGKCR 241

Qy    299 ------EETLTCIICQDLLHDCVSLQPCMHTFCA----------------ACYSGWMERSSLC 339
                     | ||||||||||||||||||||||                ||||||||||
Db    242 VVLLGDGEAQNYLPLDDTFYHSLVYDNAQTLLADKGAIRVGEKYQAVVDEWME----- 296

Qy    340 PTCRCPVERICKN-HILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSF 398
          :    ||| ||   |    :|
Db    297 ----PADREAKEAKVLAEKVKKEEEEAKLKAEEAEEDTENDGL--TIAEDEEMPELKNEE 349

Qy    399 SDEEGSSEDLLELSDVDSESS----------------DISQPYVVCRQCPEYRR--QAAQ 440
          : ::   : |:|||| ||||                  :  |
Db    350 KEETKVEE----EAITTDSEGEVLVWHPHHALTDRDIDQYMIVARSVGLFARAIDGAS 406

Qy    441 PPHCP 445
          ||  |
Db    407 APKLP 411
```

RESULT 14
T04938
hypothetical protein F7J7.10 - Arabidopsis thaliana (fragment)
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 02-Sep-2000
C;Accession: T04938
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, July 1998
A;Reference number: Z15391
A;Accession: T04938
A;Molecule type: DNA
A;Residues: 1-1331 <BEV>
A;Cross-references: EMBL:ALO21960
A;Experimental source: cultivar Columbia; BAC clone F7J7
C;Genetics:
A;Map position: 4
A;Note: intron positions not resolved
A;Note: F7J7.10
C;Superfamily: RING finger homology
F;377-424/Domain: RING finger homology <RRN>

```
Query Match             3.8%;    Score 136.5;  DB 2;  Length 1331;
Best Local Similarity  24.6%;    Pred. No. 0.98;
Matches   49;  Conservative   30;  Mismatches   69;  Indels  51;  Gaps   7;

Qy    272 LLVAQPRRNAQTVHEDVRAAAG------------------------KPDKMEET------- 301
          ||:|| ||||||||||||                              |||||||
Db    319 MLISTPERSPITVVKNLRVCADCHLAIKLVSKVYNREIVVRDRSRSKVADTSHLERMGRE 378

Qy    302 LTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVER--ICKNHILNNLV 359
          | |||||||||||||||||||||||||||||||||||||||      :||:||||||||
Db    379 LKCPICLSLYNSAVSLS-CNHVFCNACIVKSMKMDATCPVCKIPYHRREIRGAPHMDSLV 437

Qy    360 EAY---------LIQHPDKSRSEEDVQSMDARNKITQDMLQP---KVRRSFSDEEGSS 405
```

```
Db    438 SIYKNMEDASGIKLFVSQNNPSPSDKEKQVRDASVEKASDKNRQGSRKGRASKRNEYGKT 497

Qy    406 EDLLELSDVDSESSDISQP 424
          ::: |||| ||| ||:||
Db    498 KEI----DVDAPGIVMKP 512
```

RESULT 15
A85240
hypothetical protein AT4g21070 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C;Accession: A85240
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: A85240
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1495 <STO>
A;Cross-references: GB:NC_001268; NID:g7268904; PIDN:CAB79107.1; GSPDB:GN00140
C;Genetics:
C;Gene: AT4g21070
A;Map position: 4

```
Query Match             3.8%;    Score 136.5;  DB 2;  Length 1495;
Best Local Similarity  24.6%;    Pred. No. 1.1;
Matches   49;  Conservative   30;  Mismatches   69;  Indels  51;  Gaps   7;

Qy    272 LLVAQPRRNAQTVHEDVRAAAG------------------------KPDKMEET------- 301
          ||:|| ||||||||||||                              |||||||
Db    536 MLISTPERSPITVVKNLRVCADCHLAIKLVSKVYNREIVVRDRSRSKMADTSHLERMGRE 595

Qy    302 LTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVER--ICKNHILNNLV 359
          | |||||||||||||||||||||||||||||||||||||||      :||:||||||||
Db    596 LKCPICLSLYNSAVSLS-CNHVFCNACIVKSMKMDATCPVCKIPYHRREIRGAPHMDSLV 654

Qy    360 EAY---------LIQHPDKSRSEEDVQSMDARNKITQDMLQP---KVRRSFSDEEGSS 405
          |||          |||||||||||||||||||||||||||    |||||||||||||
Db    655 SIYKNMEDASGIKLFVSQNNPSPSDKEKQVRDASVEKASDKNRQGSRKGRASKRNEYGKT 714

Qy    406 EDLLELSDVDSESSDISQP 424
          ::: |||| ||| ||:||
Db    715 KEI----DVDAPGIVMKP 729
```

Search completed: May 7, 2004, 14:50:06
Job time : 40.4116 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    May  7, 2004, 14:37:32 ; Search time 22.4372 Seconds
                                   (without alignments)
                            1540.951 Million cell updates/sec

Title:          US-10-048-046-2
Perfect score:  3585
Sequence:       1 MERPEEGKQSPPPQPWGRLL.........VKAHHAMKFNHICEQTRFKN 664

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:        141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    SwissProt_42:*

       Pred. No. is the number of results predicted by chance to have a
       score greater than or equal to the score of the result being printed,
       and is derived by analysis of the total score distribution.

                              SUMMARIES
              *
Result         Query
No.   Score   Match Length DB  ID                 Description
----  -----   ----- ------ --  --                 -----------
 1    203.5    5.7    485   1   RNF8_HUMAN         O76064 homo sapien
 2    197      5.5    488   1   RNF8_MOUSE         Q8vc56 mus musculu
 3    168.5    4.7    467   1   ICP0_VZVD          P09309 varicella-z
 4    145      4.0   1603   1   PSC_DROME          P35820 drosophila
 5    139      3.9    775   1   ICP0_HSV11         P08393 herpes simp
 6    139      3.9    825   1   ICP0_HSV2H         P28284 herpes simp
 7    138      3.8    387   1   RH1B_SCHPO         O74747 schizosacch
 8    134.5    3.8    551   1   RN27_HUMAN         Q9brr9 homo sapien
 9    134      3.7    624   1   RN12_HUMAN         Q9nvw2 homo sapien
10    132.5    3.7   1300   1   SAL3_HUMAN         Q9bxa9 homo sapien
11    132.5    3.7   2414   1   P300_HUMAN         Q09472 homo sapien
12    131.5    3.7    600   1   RN12_MOUSE         Q9wtv7 mus musculu
13    131      3.7   1849   1   IGA4_HABIN         P45386 haemophilus
14    130.5    3.6    551   1   RN27_MOUSE         Q99pj2 mus musculu
15    129      3.6   1341   1   ACIN_HUMAN         Q9ukv3 homo sapien
16    128.5    3.6    933   1   PRGR_HUMAN         P06401 homo sapien
17    128      3.6    407   1   RPF2_HUMAN         O60858 homo sapien
18    128      3.6   1338   1   ACIN_MOUSE         Q9jix8 mus musculu
19    128      3.6   2003   1   NTC4_HUMAN         Q99466 homo sapien
20    127      3.5   2715   1   MLL4_HUMAN         Q9umn6 homo sapien
21    126.5    3.5    676   1   ICP0_HSVBJ         P29128 bovine herp
22    126      3.5    744   1   TRM3_HUMAN         O75382 homo sapien
23    125.5    3.5    423   1   MKR2_SERQU         Q9dd48 seriola qui
24    125.5    3.5    989   1   SERA_PLAFG         P13823 plasmodium
25    125      3.5    630   1   Z147_HUMAN         Q14258 homo sapien
26    125      3.5    744   1   TRM3_RAT           O70277 rattus norv
27    125      3.5    777   1   BAR1_HUMAN         Q99728 homo sapien
28    125      3.5    794   1   Z151_MOUSE         Q60821 mus musculu
29    124.5    3.5    705   1   ICAL_BOVIN         P20811 bos taurus
30    124.5    3.5   1127   1   TFIG_HUMAN         Q9upn9 homo sapien
31    124.5    3.5   3866   1   HRX_MOUSE          P55200 mus musculu
32    124      3.5    676   1   ICP0_HSVBK         P29836 bovine herp
33    124      3.5   1712   1   LTB1_RAT           Q00918 rattus norv
34    124      3.5    124   1   APC_RAT            P70478 rattus norv
35    123.5    3.4    892   1   ATX7_HUMAN         O15265 homo sapien
36    123.5    3.4   1233   1   M4K4_MOUSE         P97820 mus musculu
37    123      3.4    605   1   PJA1_MOUSE         O55176 mus musculu
38    123      3.4    898   1   C1Z1_HUMAN         Q9ulv3 homo sapien
39    123      3.4   1007   1   SAL2_HUMAN         Q9y467 homo sapien
40    122      3.4    433   1   TIG_CHLMU          Q9p119 chlamydia m
41    122      3.4    501   1   UVS2_NEUCR         P33288 neurospora
42    122      3.4    507   1   MKR3_HUMAN         Q13064 homo sapien
43    122      3.4    605   1   CORO_CABEL         Q21624 caenorhabdi
44    122      3.4   1004   1   SAL2_MOUSE         Q9qx96 mus musculu
45    122      3.4   1073   1   PVDA_PLAKN         P22545 plasmodium
```

                                         ALIGNMENTS

RESULT 1
RNF8_HUMAN
ID   RNF8_HUMAN   STANDARD;   PRT;   485 AA.
AC   O76064;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RING finger protein 8.
GN   RNF8 OR KIAA0646.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Seki N., Yamauchi M., Saito T.;
RT   "Isolation and chromosomal assignment of the gene for a novel zinc
RT   finger protein.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Tracey A.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length

```
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SIMILARITY: Contains 1 FHA domain.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   ---------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------
DR   EMBL; AB012770; BAA33557.1; -.
DR   EMBL; AB014546; BAA31621.1; -.
DR   EMBL; AL096712; CAB75689.1; -.
DR   EMBL; BC007517; AAH07517.1; -.
DR   Genew; HGNC:10071; RNF8.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Zinc-finger.
FT   DOMAIN        38     92       FHA.
FT   DOMAIN       276    345       GLN-RICH.
FT   ZN_FING      403    441       RING-TYPE.
SQ   SEQUENCE   485 AA;  55517 MW;  54650B2FFC9948B1 CRC64;

Query Match             5.7%;  Score 203.5;  DB 1;  Length 485;
Best Local Similarity   22.5%;  Pred. No. 2.2e-05;
Matches  102;  Conservative  60;  Mismatches 170;  Indels 121;  Gaps  17;

Qy    19  LLRLGAEEGEPHVLLRKR-EWTIGRRRGCDLSFPSN--KLVSGDHCRIVDEKSGGVTL    74
          || |      ::   :|:         :||  :||       :|| :||||||||||||
Db    20  LRRVGMSAG--WLLLEDGCEVTVGRGFGVTYQLVSKICPLMISRNHC-VLKQNPBGGQWTI  76

Qy    75  EDT-STSGTVINKLKVVKKQTCPLQTGDVIYL----------VYRKNEPEHNVAYLYESL  123
          :|: |:    :. :::  ::         ||:           :||
Db    77  MDNKSLNGVWLNRARLBEPLRVYSIHQGDYIQLGVPLENKENAEYEYEVTEEDWETIYPCL  136

Qy   124  SEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQP------  177
          | |        |::||
Db   137  SPK---NDQMIEKOKE-LRTKRKFSLDELAGPGAEGPSNLKSKINKVSCESGQPVKSQGK  192

Qy   178  ----STSTSDLFPTASA--SSTEPSPA----------GRERSSSCGSGGGGI------  213
             ::|:|  |||
Db   193  GEVASTPSDNLDPKLTALEPSKTTGAPIYPGFPKVTEVHEQKASNSSASQRSLQMFKVT  252

Qy   214  ------------------SPKGGSGPSVAS-----DEVSSFASALPDRKTASFSS  244
                            | |     ||
Db   253  MSRILRLKIQMQEKHEAVMNVKKQTQKGNSKKVVQMEQELQDLQSQLCAEQAQQQARVEQ  312

Qy   245  LEPQDQDBLEPVKKKMRGDGDLDLNGQL------------------LVAQPRRN  280
          || |    || ||
Db   313  LEKTFQEEECHLQGLEIAQGEKDLKQQLAQALQEHWALMEELNRSKKDFEAIIQAKNKEL  372

Qy   281  AQTVHEDVRAAAGKP--------DKMEETLTCIICCDLLHDCVSLQPCMHTFCAACYSGWM  333
                           :|       | ||  : |||| ::|:  :| | || || ||
Db   373  EQTKEKEKMQAQKEVLSHMNDVLENELQCIICSEYFIEAVTLN-CAHSFCSYCINEWM  431

Qy   334  ERSSLCPTCRCPVER------ICKQNHILNNL  358
           : | |  |    |:       ||:|| | |:
Db   432  KRKIEBCPICRKDIKSKTYSLVLDNCINKMVNNL  464

RESULT 2
RNF8_MOUSE
```

```
ID   RNF8_MOUSE     STANDARD;     PRT;   488 AA.
AC   Q8VC56;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RING finger protein 8.
GN   RNF8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SIMILARITY: Contains 1 FHA domain.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   ---------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------
DR   EMBL; BC021778; AAH21778.1; -.
DR   MGD; MGI:1929069; RNF8.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Zinc-finger.
FT   DOMAIN        38     92       FHA.
FT   DOMAIN       279    345       GLN-RICH.
FT   ZN_FING      406    444       RING-TYPE.
SQ   SEQUENCE   488 AA;  55516 MW;  428242204EBC44A1 CRC64;

Query Match             5.5%;  Score 197;  DB 1;  Length 488;
Best Local Similarity   21.0%;  Pred. No. 5.5e-05;
Matches   98;  Conservative  55;  Mismatches 173;  Indels 140;  Gaps  17;

Qy    27  GEPHVLL----RKREWTIGRRRGCD---------------------------LSFPSNKLVS  57
              ::       :     :                               :  ::  :
Db     2  GEPDPLVSGQLAARRSWCL-RRLGMDCEWLQLEAGTEVTIGRGLSVTYQLISKVCPLMIS  60

Qy    58  GDHCRIVDEKSGGVTLEDT-STSGTVINKLKVVKKQTCPLQTGDVIYLVVYRKNEPEHNV  116
```

```
Db   61 RSHC-VLKQNPEGQWTIMDNKSLNGVWLNRERLAPLQGYCIRKGDHIQLGV-PLESRETA 118
         ||  :: ||  | :  |||: : :||  :| : ||: . ::| .:|||
Qy  117 AYLYESLSE------------------------KQGMTQESFEA-NKENV----------- 141
         :|:|:                           .:|:   :.:
Db  119 EYEYRVIEEDWESLAPCLAPKNDQRMEKHKGSRTKRKFSSPGLENLPAEGSSDLRCPLAN 178

Qy  142 ------FHGTKDTSGAGAG---------RGADPRVPPSSPAT--------Q 169
                                            :
Db  179 VASKPIEPEKIHGKGDASSQSLGCLCPGLTSLKASERAAGPHACSALPKVLELSCPKKQK 238

Qy  170 VCFEERQPSTSTSDLFPTASASSTEPSPAGRERSSSCGSGGGGISPKGSGPSVASDE--- 226
            .: .::.:|
Db  239 AC--RPSASQNSLELFKVTMSRMLKTQMEKQIAV-LNVKRQTRKGSSKKIVRMEKSL   295

Qy  227 --VSSFASALPDRKTASFSSLEPQDQEDLEPVKKMRGGDLDLNGQLLVAQPRRNA---   281
            ||  .:|  :.|  ||:  ||  |: :|  .:  |:|:  :||
Db  296 RNLQSQLYAEQAQQQARVEQLEKTFQEEAHYLQGLEKEQGECDLKQQLVQALQEHQALME 355

Qy  282 -----------QTVHEDVRAAAGKP---------DKMEETLJCIICQDLLHD       313
                       ||  :    |||       :||  | | |||:
Db  356 ELNCSKKDFEKIIQAKNKELEQTKEEKDKVQAQKEEVLSHMNDLLENELQCIICSEVFIE 415

Qy  314 CVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLV 359
         || .||||||| ||||: :  .||||||.|  :||: ::|: |
Db  416 AVTLN-CAHSFCSFCINEWMKRKVECPICRKDIESRTNSIVLDNCI 460
```

RESULT 3
ICP0_VZVD
```
ID   ICP0_VZVD    STANDARD;    PRT;    467 AA.
AC   P09309;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Trans-acting transcriptional protein ICP0.
GN   61.
OS   Varicella-zoster virus (strain Dumas) (VZV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=10338;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86306657; PubMed=3018124;
RA   Davison A.J., Scott J.E.;
RT   "The complete DNA sequence of varicella-zoster virus.";
RL   J. Gen. Virol. 67:1759-1816(1986).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=93059681; PubMed=1366099;
RA   Moriuchi H., Moriuchi M., Smith H.A., Straus S.E., Cohen J.I.;
RT   "Varicella-zoster virus open reading frame 61 protein is functionally
RT   homologous to herpes simplex virus type 1 ICP0.";
RL   J. Virol. 66:7303-7308(1992).
CC   -!- FUNCTION: REPRESSE THE EXPRESSION OF VIRAL IE, EARLY, AND LATE
CC       GENE PROMOTERS.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.
```
```
DR   EMBL; X04370; CAA27944.1; -.
DR   PIR; I27225; WZBE61.
DR   HSSP; P28990; 1CHC.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS00089; ZF_RING_2; 1.
KW   Transcription regulation; Trans-acting factor; Repressor; Zinc-finger;
KW   DNA-binding.
FT   ZN_FING    19     58    RING-TYPE.
SQ   SEQUENCE   467 AA;   50916 MW;   25EFA6977ZA6994C CRC64;
```

```
Query Match      4.7%;   Score 168.5;   DB 1;   Length 467;
Best Local Similarity   26.3%;   Pred. No. 0.0028;
Matches  64;   Conservative  20;   Mismatches 106;   Indels 53;   Gaps  10;
```

```
Qy  287 DVRAAAGKPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPV 346
         |:  |:  || :  |||| |  :| ||||||| |||| :  ||||||.|.|||
Db    2 DTILAGGSGTSDASDNTCTICMSTVSDLGKTMPCLHDFCFVCIRAWTSTSVQCPLCRCPV 61

Qy  347 ERICKNHILNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSE  406
         | |  : :    :   ::|          |:
Db   62 QSILHKIVSDTSYKEYEV-HP----SDDDGFS---------EPSFEDSIDILPGDVI   104

Qy  407 DLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPBGEPGAPQALGDAPSTSVS 466
         :||  :  : .  ||  ::  :
Db  105 DLLPPSPGPSRES-IQQP--------TSRSSREPIQSP----NPGPLQSSARBP-TAES  149

Qy  467 LTTAVQDYVCP-LQGSHALCTCC------------EREQDPRV                  505
         :|::  |
Db  150 PSDSQQDSIQPPTRDSSPGVTKTCSTASFLRKVFFKDQPAVRSATPVVYGSIESAQQPRT 209

Qy  506 APQ 508
         |
Db  210 GGGQ 212
```

RESULT 4
PSC_DROME
```
ID   PSC_DROME    STANDARD;    PRT;    1603 AA.
AC   P35820;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Posterior sex combs protein.
GN   PSC.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92018190; PubMed=1833647;
RA   Brunk B.P., Martin E.C., Sharp E., Adler P.N.;
RT   "Drosophila genes Posterior Sex Combs and Suppressor two of zeste
RT   encode proteins with homology to the murine bmi-1 oncogene.";
RL   Nature 353:351-353(1991).
CC   -!- FUNCTION: The Polycomb group (Pc-G) genes are needed to maintain
CC       expression patterns of the homeotic selector genes of the
CC       antennapedia (Antp-C) and bithorax (BX-C) complexes, and hence for
CC       the maintenance of segmental determination.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
```
```
DR   EMBL; X59275; CAA41965.1; -.
DR   PIR; S17983; S17983.
DR   FlyBase; FBgn0005624; Psc.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
```

```
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS00089; ZF_RING_2; 1.
KW   Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
FT   DOMAIN        47     53       POLY-THR.
FT   DOMAIN        83     88       POLY-THR.
FT   DOMAIN        91     98       POLY-THR.
FT   DOMAIN       145    152       POLY-THR.
FT   DOMAIN       184    202       POLY-SER.
FT   ZN_FING      265    304       RING-TYPE.
FT   DOMAIN       642    651       POLY-SER.
FT   DOMAIN      1066   1069       POLY-GLY.
FT   DOMAIN      1185   1189       POLY-PRO.
FT   DOMAIN      1214   1217       POLY-PRO.
FT   DOMAIN      1391   1396       POLY-PRO.
FT   DOMAIN      1458   1461       POLY-ALA.
FT   DOMAIN      1517   1520       POLY-GLY.
SQ   SEQUENCE   1603 AA;  169999 MW;  77024F4097736473 CRC64;

Query Match            4.0%;  Score 145;  DB 1;  Length 1603;
Best Local Similarity  20.3%;  Pred. No. 0.36;  Indels  56;  Gaps  12;
Matches   84;  Conservative  70;  Mismatches  203;

Qy    67  EKSGQVTLEDTSTSGTV---INKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLYES 122
                               ::    :    :::
Db    17  KQTAEATPEATATTTMAHTQQKSQLSTLAKTTTTTTATNKAAKSVVSNANSSGNNSGKKLA 76

Qy   123  LSEKQGMTQESFEANKENVFHGTKDTSGAGARGADPRVPPSSPATQVCF--------E 173
           ||: :                                               ::
Db    77  LSQSQKTTTTTTTPFTT----TTTTTAAEATTNAD-KMQKQQQLKQQLFAACSIKVKSE 130

Qy   174  EPQPSTSTSDLFPTASASSTEPSPAGRERSSSCCGSGGGGISPKGSGPSVASDEVSSFASA 233
           ||        | |                |    :|  :: :::|::::
Db   131  NALATTATAAL-AAATTTTTTATPALATGKAAKTILENGIKKESTPPAVESVEASSSSSS 189

Qy   234  LPDRKTASFSSSLEPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAG 293
                :              :: :              :         ::
Db   190  SSSSSSSSSSSWSTTRRATSEDASSNGGASADEEKSEEDPTAAVAASSTATTTSDLATTS 249

Qy   294  KPDKMEET-----LTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVER 348
                        :||:||||:|  |:||| |||:|| |||:  : : :|||||| || ::
Db   250  RPRPVLLTAVNPHIICHLCQGYLINATTIVECLHSFCHSCLINHLRKERFCPRCEMVINN 309

Qy   349  ICKN--------HILNNLVEAYLIQHPD--KSRSEE-DVQSMDARNKITQDMLQPK 393
                        |  :::  :::                  : ::
Db   310  AKPNIKSDTTLQAIVYKLVPGLYERELMRKRAFYKDRPEEAALATPEQRGDDTEHLI--- 366

Qy   394  VRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPA 446
                :        ::  :  ||: : :  :
Db   367  --FSPSDDMSLSLEYAELGELKTD---SEPELVDTLRPRYLQ-------CPA 406

RESULT 5
ICP0_HSV11
ID   ICP0_HSV11   STANDARD;   PRT;   775 AA.
AC   P08393;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Trans-acting transcriptional protein ICP0 (Immediate-early protein
DE   IE110) (VMW110) (Alpha-0 protein).
GN   ICP0 OR IE110.
OS   Herpes simplex virus (type 1 / strain 17).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10299;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88274327; PubMed=2839594;
RA   McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA   McNab D., Perry L.J., Scott J.E., Taylor P.;
RT   "The complete DNA sequence of the long unique region in the genome of
RT   herpes simplex virus type 1.";
RL   J. Gen. Virol. 69:1531-1574(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87059760; PubMed=3023529;
RA   Perry L.J., Rixon F.J., Everett R.D., Frame M.C., McGeoch D.J.;
RT   "Characterization of the IE110 gene of herpes simplex virus type 1.";
RL   J. Gen. Virol. 67:2365-2380(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89036163; PubMed=2846760;
RA   Perry L.J., McGeoch D.J.;
RT   "The DNA sequences of the long repeat region and adjoining parts of
RT   the long unique region in the genome of herpes simplex virus type
RT   1.";
RL   J. Gen. Virol. 69:2831-2846(1988).
CC   -!- FUNCTION: ICP0 IS A GENERAL TRANSACTIVATOR OF ALL THREE CLASSES OF
CC       HSV GENES WHICH FUNCTIONS SYNERGISTICALLY WITH ICP4. OFTEN TERMED
CC       AS A PROMISCUOUS TRANSACTIVATOR. MAY PLAY A ROLE IN THE BALANCE
CC       BETWEEN THE LATENT AND LYTIC STATES, REACTIVATING LATENT HSV. ICP0
CC       MIGRATES TO THE ND10 NUCLEAR BODIES IN THE EARLY TIMES OF
CC       INFECTION AND DISRUPTS THEM. BINDS TO HAUSP AND MAY MODIFY ITS
CC       DEUBIQUITINATING SUBSTRATE SPECIFICITY OR ACTIVITY ON VIRAL OR
CC       CELLULAR TARGETS, LEADING TO AN INCREASED OR DECREASED STABILITY
CC       OF THESE PROTEINS.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.
```

```
DR   EMBL; X14112; CAA32336.1; -.
DR   EMBL; X14112; CAA32293.1; -.
DR   EMBL; X04614; CAA28285.1; -.
DR   PIR; A29152; BDBl1.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS00089; ZF_RING_2; 1.
KW   Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
KW   DNA-binding; Early protein.
FT   ZN_FING      116    157       RING-TYPE.
FT   DOMAIN       233    243       POLY-ASP.
FT   DOMAIN       305    308       POLY-GLY.
FT   DOMAIN       558    568       POLY-SER.
SQ   SEQUENCE   775 AA;  78456 MW;  DF38A1C539DAB15C CRC64;

Query Match            3.9%;  Score 139;  DB 1;  Length 775;
Best Local Similarity  22.6%;  Pred. No. 0.33;
Matches   80;  Conservative  34;  Mismatches  142;  Indels  98;  Gaps  16;

Qy   156  GADPRVPPSSPATQVCFEEPQPST----STSDLFPTASASSTEPSPAGRERSSSCCGSGSGG 211
            |:  ||   :                  :||||||||||||||||||||||||||:|||| :
Db     6  GASTRRPBGRPQ----REPAPDVWVFPCDRDL-PDSSDSEAETEVGGR---------- 48

Qy   212  GISPKGSGPSVASDEVSSFASALPDRKTASFSSSLEPQDQEDLEPVKKKMRGDGDLDLNGQ 271
                                      :                 ::
Db    49  ------GDADHHDD----------DSASEADSTDTBLFETGLLGPQGVD-GGA 84

Qy   272  LLVAQPRRNAQTVHEDVRAAAGKPDK---MEETLLTCIICQDLL-HDCVSLQPCMHTFC 325
            :  |         :              ||||:||||||||| |||  | ||||:||
Db    85  VSGGSPPR-----EEDPGSCGGAPPREDGGSDEGDCAVCTEIAPHLRCDTFPCMHRFC 139

Qy   326  AACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSR--SEEDVQSMDARN 383
            :| :  ||| :||| |  |   :     :  :::
Db   140  IPCMKTWMQLRNTCPLCNAKLVLIVGVTPSGSFSTIPIVNDPQTRMEAEEAVRAGTAVD 199

Qy   384  KI---------TQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYV 426
```

```
Db   200 FIWTGNQRFAPRYLTLGGHTVRALSPTHEPTTDED--------DDDLDDADYVPP-- 247
             |||:                      |:|              :||:
Qy   427 VCRQCPEY--RRQAAQPP------HCPAPEGEPGAPQA--LGDAPSTSVSLT 469
             :|:||                   ::|              :|:|
Db   248 APRRTPRAPPRRGAAAPPVTGGASHAAPQPAAARTAPPSAPIGPHGSSNTNTTT 301
```

```
RESULT 6
ICP0_HSV2H
ID   ICP0_HSV2H    STANDARD;      PRT;    825 AA.
AC   P28284;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Trans-acting transcriptional protein ICP0 (VMW118 protein).
GN   RL2.
OS   Herpes simplex virus (type 2 / strain HG52).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus;
OX   NCBI_TaxID=10315;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92113549; PubMed=1662697;
RA   McGeoch D.J., Cunningham C., McIntyre G., Dolan A.;
RT   "Comparative sequence analysis of the long repeat regions and
RT   adjoining parts of the long unique regions in the genomes of herpes
RT   simplex viruses types 1 and 2.";
RL   J. Gen. Virol. 72:3057-3075(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Dolan A.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
-!- SIMILARITY: Contains 1 RING-type zinc finger.
-!- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.
-----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
-----------------------------------------------------------------------
DR   EMBL; D10471; BAA23427.1; -.
DR   EMBL; Z86099; CAB06760.1; -.
DR   PIR; JQ1501; EDDEXD.
DR   HSSP; P28990; 1CHC.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
KW   DNA-binding.
FT   DOMAIN     120    123       POLY-GLY.
FT   ZN_FING    126    167       RING-TYPE.
FT   DOMAIN     266    271       POLY-GLY.
FT   DOMAIN     292    295       POLY-SER.
FT   DOMAIN     342    345       POLY-ALA.
FT   DOMAIN     386    389       POLY-SER.
FT   DOMAIN     395    400       POLY-GLY.
FT   DOMAIN     425    428       POLY-ALA.
FT   DOMAIN     590    627       POLY-SER.
SQ   SEQUENCE  825 AA;  81986 MW;  5CEB15B5B553A274 CRC64;

Query Match             3.9%;  Score 139;  DB 1;  Length 825;
Best Local Similarity  22.7%;  Pred. No. 0.36;
Matches    82;  Conservative   33;  Mismatches  127;  Indels  120;  Gaps   19;

Qy   152 GAGRGADP--RVPPSSPATQVCFEFQPSTSTSDLPPTASASTEPSPAG-------RB 201
             |||:                                      :||:||   |:
Db     6 GTSSRADPGPERPPRQTPGTQPA--APHAWGMLNDMQWLASDSEETEVGISDDDLHRD 63
```

```
Db   202 RSSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTASFSSLEPQDEDLEPVKKMR 261
             ||||                  :|                    ||:
Qy    64 STSEAG----------------STDTEMFEAGLMDAAT---------PPARPAERQG 96
             :|                    :|:                    ||:|||::
Db   262 GDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEETLTCIICQDLLHDCVSLQ--P 319
             ::                    :|                    :||: ||:
Qy    97 SPTPADACGSC------------GGGPVGEEEABAGGGGD----VCAVCTDIAPPLRCQSFP 143
             :||                   :|                    |||||
Qy   320 CMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVEAYLIQHPDKSREEDVQSM 379
             ||||                  :|                    :||||||
Db   144 CLHPFCIPCMKTWIPLRNTCPLCNTPV-------------------AYLIVGVTASSGFSTIPIV 189
             :|||                  :|
Qy   380 -DARNKITQDML-----------------QPKVR-RSFS--------DEBGSSEDLLEL 411
             :|||                  :||                    |:
Db   190 NDPRTRVEAEAAVRAGTAVDFIWTGNPRTAPRSLSLGGHTVRALSPTTPPWPGTDDEDDDL 249
             ::||                  |||                    ||||:
Qy   412 SDVDSESSDISQPYV--VCRQCPE------------YRRQAAQPPHC-PAPEGEPGAPQ--ALGD 459
             :|||                  |||                    ||||
Db   250 ADVD-----------------YVPPAPRRAPRRGGGAGATRGTSQPAATRPAP--PGAPRSSSSGG 297

Qy   460 AP 461
             ||
Db   298 AP 299
```

```
RESULT 7
RH18_SCHPO
ID   RH18_SCHPO    STANDARD;      PRT;    387 AA.
AC   O74747;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Postreplication repair protein rhp18 (RAD18 homolog).
GN   RHP18 OR SPBC1734.06.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A. AND FUNCTION.
RX   MEDLINE=21559218; PubMed=11702950;
RA   Kitamura K., Katayama S., Dhut S., Sato M., Watanabe Y., Yamamoto M.,
RA   Toda T.;
RT   "Phosphorylation of Mei2 and Ste11 by Pat1 kinase inhibits sexual
RT   differentiation via ubiquitin proteolysis and 14-3-3 protein in
RT   fission yeast.";
RL   Dev. Cell 1:389-399(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schafer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
```

RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Involved in postreplication repair of UV-damaged DNA.
CC   Postreplication repair functions in gap-filling of a daughter
CC   strand on replication of damaged DNA. Has ssDNA binding activity
CC   (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- SIMILARITY: Contains 1 SAP domain.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AB079544; BAB84669.1; -.
DR   EMBL; AL031856; CAA21300.1; -.
DR   PIR; T39653; T39653.
DR   HSSP; P15919; 1RMD.
DR   GeneDB_SPombe; SPBC1734.06; Rad18.
DR   InterPro; IPR004580; SAP.
DR   InterPro; IPR003034; SAP.
DR   InterPro; IPR006642; Znf_Rad18.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF02037; SAP; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   SMART; SM00513; SAP; 1.
DR   SMART; SM00734; ZnF_Rad18; 1.
DR   TIGRFAMs; TIGR00599; rad18; 1.
DR   PROSITE; PS50800; SAP; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Zinc-finger; DNA-binding; DNA damage; DNA repair; Nuclear protein.
FT   ZN_FING   29   67   RING-TYPE.
FT   DOMAIN   240   274   SAP.
FT   DOMAIN   181   190   POLY-SER.
SQ   SEQUENCE   387 AA; 43437 MW; 1E8140F8B645299B CRC64;

Query Match   3.8%; Score 138; DB 1; Length 387;
Best Local Similarity   25.2%; Pred. No. 0.16; Indels 14; Gaps   6;
Matches 38; Conservative 28; Mismatches 71; Indels 14; Gaps   6;

Qy   298 MEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVE--RICKONHIL 355
              :: :||:|| ||| :| || || |||:   ::|| |: |||||||||
Db   23 LDSSLRCLICHEYPR-APLITSCCHTFCSFCIRDYLREHPMCPACRAPEQBSRLRKNTIL 81

Qy   356 NNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVD 415
              ::|:|:| |: |:|: ||| ||||| | :::|| |:
Db   82 EEILESFKVIRPTL-----FEFLKVEN-VPKPVLQAPETVIAQDSASGDEEWED--DLA 132

Qy   416 SESSDISQPYVVCRQCPEYRRQAAQPPHCPA 446
              |:||::| :||:   :|: ||||
Db   133 SNSSPASIAKKTSRDSKKRKRE--DLVHCPA 161

RESULT 8
RN27_HUMAN
ID   RN27_HUMAN   STANDARD;   PRT;   551 AA.
AC   Q9BZR9; Q9C028;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RING finger protein 27 (Glioblastoma-expressed ring finger protein)
DE   (Tripartite motif-containing protein 8).
GN   TRIM8 OR RNF27 OR GERP.

OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20568703; PubMed=11118312;
RA   Vincent S.R., Kwasnicka D.A., Fretier P.;
RT   "A novel RING finger-B box-coiled-coil protein, GERP.";
RL   Biochem. Biophys. Res. Commun. 279:482-486(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Neuron;
RX   MEDLINE=21231161; PubMed=11331580;
RA   Reymond A., Meroni G., Fantozzi A., Merla G., Cairo S., Luzi L.,
RA   Riganelli D., Zanaria E., Messali S., Cainarca S., Guffanti A.,
RA   Minucci S., Pelicci P.G., Ballabio A.;
RT   "The tripartite motif family identifies cell compartments.";
RL   EMBO J. 20:2140-2151(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBUNIT: Homodimer.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Widely expressed.
CC   -!- DOMAIN: The coiled coil domain is required for homodimerization.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- SIMILARITY: Contains 2 B box-type zinc fingers.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AF281046; AAG53087.1; -.
DR   EMBL; AF220034; AAG53488.1; -.
DR   EMBL; BC021925; AAH21925.1; -.
DR   PIR; JC7562; JC7562.
DR   Genew; HGNC:15579; TRIM8.
DR   MIM; 606125; -.
DR   GO; GO:0008270; F:zinc ion binding; NAS.
DR   InterPro; IPR000315; Znf_Bbox.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00643; zf-B_box; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50119; ZF_BBOX; FALSE_NEG.
DR   PROSITE; PS00518; ZF_RING_1; 1.

```
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Nuclear protein; Coiled coil; Zinc-finger; Repeat.
FT   ZN_FING       15     56       RING-TYPE.
FT   ZN_FING       92    132       B BOX-TYPE 1.
FT   ZN_FING      140    182       B BOX-TYPE 2.
FT   DOMAIN       181    249       COILED COIL (POTENTIAL).
FT   CONFLICT     174    174       H -> R (IN REF. 1).
SQ   SEQUENCE     551 AA;   61489 MW;  1FEF89029B89BACB CRC64;

Query Match             3.8%;  Score 134.5;  DB 1;  Length 551;
Best Local Similarity  23.8%;  Pred. No. 0.41;  Indels  87;  Gaps  11;
Matches  58;  Conservative  20;  Mismatches  79;

Qy   299  EETLTCIICQDLLHDCVSLQPCMHTFCAACY-SGWMERSSL--CPTCRCPVER---ICKN  352
              |||   |  |      |       || |:  |         ||  ||  ||  ||
Db    10  EEBLICPICLHVFVEPVQL-PCKHNFCRGCIGAWAKDSGLVRCPECNQAYNQKPGLEKN  68

Qy   353  HILNNLVEAYLIQHPDKS--------------------RSEEDVQSMDARNKITQDML  390
              | |:||  | :|:                       :  :  :  :  |
Db    69  LKLTNIVEKFNALHVEKPPAALHCVFCRRGPPLPAQKVCLRCEAPC----CQSHVQTHLQ  124

Qy   391  QPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGB  450
              |    :                                  :||
Db   125  QPSTARGHL-------LVEADDVRAWS----------------CPQH--NAYRLYHCEAEQ--  160

Qy   451  PGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCC-------------FQPMPDRRAB  498
                                                        :||:
Db   161  -------------------VAVCQYCCYYSGAHQGHSVCVDVEIRRNEIRKMLMKQQDRLEB  202

Qy   499  REQD  502
              ||||
Db   203  REQD  206
```

```
RESULT 9
RN12_HUMAN
ID   RN12_HUMAN    STANDARD;    PRT;   624 AA.
AC   Q9NVW2; Q9Y598;
DT   16-OCT-2001  (Rel. 40, Created)
DT   16-OCT-2001  (Rel. 40, Last sequence update)
DT   28-FEB-2003  (Rel. 41, Last annotation update)
DE   RING finger protein 12 (LIM domain interacting RING finger protein)
DE   (RING finger protein LIM domain-binding protein) (R-LIM) (NY-REN-43 antigen).
GN   RNF12 OR RLIM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20469411; PubMed=11013082;
RA   Ostendorff H.P., Bossenz M., Mincheva A., Copeland N.G., Gilbert D.J.,
RA   Jenkins N.A., Lichter P., Bach I.;
RT   "Functional characterization of the gene encoding RLIM, the
RT   corepressor of LIM homeodomain transcription factors.";
RL   Genomics 69:120-130(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Renal cell carcinoma;
RX   MEDLINE=99438124; PubMed=10508479;
RA   Scanlan M.J., Gordan J.D., Williamson B., Stockert E., Bander N.H.,
RA   Jongeneel C.V., Gure A.O., Jager D., Jager E., Knuth A., Chen Y.-T.,
RA   Old L.J.;
RT   "Antigens recognized by autologous antibody in patients with renal-
RT   cell carcinoma.";
RL   Int. J. Cancer 83:456-464(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
```

```
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Acts as a negative co-regulator for LIM homeodomain
CC       transcription factors. Via the recruitment of the SIN3a/histone
CC       deacetylase corepressor complex.
CC   -!- SUBUNIT: Associates with LIM/homeobox factors.
CC   -!- TISSUE SPECIFICITY: Expressed in many tissues.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC       in position 134 and 142.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ271670; CAC14228.1; -.
DR   EMBL; AF155109; AAD42875.1; ALT_FRAME.
DR   EMBL; AK001334; BAA91632.1; -.
DR   Genew; HGNC:13429; RNF12.
DR   MIM; 300379; -.
DR   GO; GO:0017053; C:transcriptional repressor complex; NAS.
DR   GO; GO:0003714; F:transcription co-repressor activity; NAS.
DR   GO; GO:0016481; P:negative regulation of transcription; NAS.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Transcription regulation; Zinc-finger.
FT   DOMAIN       422    506       SER-RICH.
FT   ZN_FING      570    611       RING-TYPE.
FT   DOMAIN       453    481       POLY-SER.
FT   DOMAIN       500    506       POLY-SER.
FT   CONFLICT     126    126       S -> C (IN REF. 1).
FT   CONFLICT     134    134       N -> D (IN REF. 3).
FT   CONFLICT     144    145       YS -> NR (IN REF. 3).
FT   CONFLICT     418    418       Y -> H (IN REF. 3).
SQ   SEQUENCE     624 AA;   68527 MW;  DE3ADE09ACACBCF8 CRC64;

Query Match             3.7%;  Score 134;  DB 1;  Length 624;
Best Local Similarity  22.3%;  Pred. No. 0.52;  Indels 106;  Gaps  9;
Matches  55;  Conservative  34;  Mismatches  52;

Qy   100  GDVILVLYRKNEPEHNVAYLYBSLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADP  159
              ||:  |||:||    :||    |::::::                           :
Db   414  GELSYFMYSDSDSEPT------GSVSNRNMERAES--------RSGRGGSGGGSSSGSSS  459

Qy   160  RVPPSSPATQVCFEEPQPSTSTSDLFPTASASTBPSPAGRERSSSCGSGGGISPKGSG  219
                |||  ||          ||||||||                      ||||||
Db   460  SSSSSSSSSSSSSSPSSSSGG--ESSETSSDLFBGSNEGSSSSGSSGARREGRHRA  516

Qy   220  PSVASDEVSSFASALPDRKTASPSSLEPDQDELBPVKKMRGDGDLDLNGQLLVAQPR-  278
              ||   :  |||            |||||||||||||
Db   517  P-VTFDE---SGSLPFLSLAQFFLLNEDDDD----------------------QPRG  547

Qy   279  RNAQTVHEDVRAAAGKPDKMEETLTCIICQDLLIHDCVSLQ--PCMHTFCAACYSGWMERS  336
                                ||||||||||||||||||||
Db   548  LTKEQIDNLAMRSFGENDALK----TCSVCITEYTEGNKLRKLPCSHYHVHCIDRWLSEN  604

Qy   337  SLCPTCR  343
              |||||||
Db   605  STCPICR  611
```

```
RESULT 10
SAL3_HUMAN
ID   SAL3_HUMAN    STANDARD;    PRT;   1300 AA.
```

```
AC   Q9BXA9; Q9UGH1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Sal-like protein 3 (Zinc finger protein SALL3) (hSALL3).
GN   SALL3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RA   Gooding R., Angelicheva D., Blechschmidt K., Swoboda K., Molnar M.,
RA   Tournev I., Kalaydjieva L.;
RT   "Exclusion of HSALL3 and refinement of the region for the CCFDN
RT   gene.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 29-1300 FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=20079154; PubMed=10610715;
RA   Kohlhase J., Hausmann S., Stojmenovic G., Dixkens C., Bink K.,
RA   Schulz-Schaeffer W., Altmann M., Engel W.;
RT   "SALL3, a new member of the human spalt-like gene family, maps to
RT   18q23.";
RL   Genomics 62:216-222(1999).
RN   [3]
RP   VARIANT LEU-593, AND REVISIONS TO 787; 797-802; 808; 1138 AND 1141.
RA   Kohlhase J.;
RL   Submitted (JUL-2002) to Swiss-Prot.
CC   -!- FUNCTION: Probable transcription factor.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC       Name=3;
CC         IsoId=Q9BXA9-1; Sequence=Displayed;
CC       Name=1;
CC         IsoId=Q9BXA9-2; Sequence=VSP_006833; Major isoform with isoform
CC         Note=Lacks two zinc finger domains. Major isoform with isoform
CC         2;
CC       Name=2;
CC         IsoId=Q9BXA9-3; Sequence=VSP_006832, VSP_006833;
CC         Note=Lacks two zinc finger domains. Major isoform with isoform
CC         1;
CC       Name=4;
CC         IsoId=Q9BXA9-4; Sequence=VSP_006832;
CC   -!- TISSUE SPECIFICITY: Widely expressed in adult with highest levels
CC       in heart. Expressed in fetal brain (in neurons of hippocampus,
CC       cortex, mediodorsal and ventrolateral thalamic nuclei, putamen,
CC       cerebellum and brainstem).
CC   -!- DEVELOPMENTAL STAGE: In fetal brain of the 24th gestational week.
CC   -!- SIMILARITY: BELONGS TO THE SAL FAMILY OF C2H2-TYPE ZINC-FINGER
CC       PROTEINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF347021; AAK18311.1; -.
DR   EMBL; AJ007421; CAB65124.1; ALT_SEQ.
DR   HSSP; P07248; 1ARE.
DR   Genew; HGNC:10527; SALL3.
DR   MIM; 605079; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 10.
DR   SMART; SM00355; ZnF_C2H2; 10.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
KW   Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
KW   Metal-binding; Repeat; Alternative splicing; Polymorphism.
```

```
FT   ZN_FING      420    442       C2H2-TYPE.
FT   ZN_FING      448    470       C2H2-TYPE.
FT   ZN_FING      679    701       C2H2-TYPE.
FT   ZN_FING      707    729       C2H2-TYPE.
FT   ZN_FING      739    761       C2H2-TYPE.
FT   ZN_FING      977    999       C2H2-TYPE.
FT   ZN_FING     1005   1027       C2H2-TYPE.
FT   ZN_FING     1113   1135       C2H2-TYPE.
FT   ZN_FING     1141   1163       C2H2-TYPE.
FT   DOMAIN       198    201       POLY-ALA.
FT   DOMAIN       213    216       POLY-GLN.
FT   DOMAIN       897    902       POLY-SER.
FT   VARSPLIC       1    133       Missing (in isoform 2 and isoform 4).
FT                                 /FTId=VSP_006832.
FT   VARSPLIC     973   1044       Missing (in isoform 1 and isoform 2).
FT                                 /FTId=VSP_006833.
FT   VARIANT      593    593       V -> L.
FT                                 /FTId=VAR_014132.
FT   CONFLICT     235    235       R -> C (IN REF. 2).
SQ   SEQUENCE   1300 AA;  135371 MW;  82CF3BDCB6D59150 CRC64;
```

```
Query Match             3.7%;  Score 132.5;  DB 1;  Length 1300;
Best Local Similarity  21.5%;  Pred. No. 1.6;
Matches  134;  Conservative  61;  Mismatches  220;  Indels  209;  Gaps  34;

Qy    127 QGMTQESFEANKENVFH--------------GTKDTSGAGARGADPRVPPSSPATQVCFE 173
          ::||::  :  : : :              ::|  :  |   :
Db    460 KGNLKVHFQRHEKEKYPHIQMNPYPVPEYLDNVPTCSGIPYGMS----LPPEKPVTTWLDS 515

Qy    174 EP-QPSTSTS--DLFPT-----------------ASASTEPSPAGRERSSSCCGSGGGGI- 213
          ||  :|:|:  :||||                  :  :
Db    516 KPVLPTVPTSVGLQLPPTVPGAHGYADSPSATPASRSPQRPSPA----SSECASLSPGLN 571

Qy    214 -----------SPKG--SGPSVASDE----------------VSSFASALPDRKTASFSSL 245
                     :|:|  ::|:|:                  ||  :  :||
Db    572 HVESGVATAESPQSLLGGPPVTKAEPVSLPCTNARAGDAPVGAQASAAPTSVDGAFTSL 631

Qy    246 -EPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEETLTC 304
           ||  ||:  |:       |||::|:|:   |:                ::
Db    632 GSPGLPAVSEQFKAQFPFGGLID---SMQTSETSKLQQLV-ENI------DKXMTDPNQC 681

Qy    305 IICQDLLHDCVSLQPCM-----HT----FCAACYSGWMERSSL------------CPTCR- 343
          ::|  ||:|:|:|:  |      :|      |
Db    682 VIC---HRVLSCQSALKMHYRTHTGERPFFKCKICGRAFTTKGNLKTHFGVHRAKPPLRV 737

Qy    344 ---CPVERICKNHILNNIVEAYLIQH-----------PDKSRSEEDVQSMDARNKITQDMLQ 391
             :||: :||  : : :  : |:|           ::|
Db    738 QHSCP---ICQKKFTNAVV---LQQHIRMEHMGGQIPNTPLPEGFQDAMD--SELAYDDKN 789

Qy    392 PKVRRSFSD--BEGSSEDLLELSDVDSESSDISQPYY-VCRQCPEYRRQAAQPP------- 442
          |  :  :|:  :|         |   ::: : :|::
Db    790 AETLSSYDDMDENSMEDDAELKDA--ATDPAKPLLSYAGSCP---------PSPPSVISSI 840

Qy    443 ----------HCPAPEGEBPGAPQALGADPASTSVSLTTAVQDYVCPLQGSHAL--- 484
                    :|:                          :|:
Db    841 AALENQMKMIDSVMSCQQLTGLKSVENGSGESDRLSNDSSSAVGDLESRSAGSPALSESS 900

Qy    485 CTCCFQPMPDR----------------------------RAEREQDPRVAPQQCAVCLQ--- 515
          :  :  | |                            : |  |:
Db    901 SSQALSPAPSNGESFRSKSPGLGAPEEPQEIPLKTERPDSPAAAPGSGGAPGRAGIKEEA 960

Qy    516 PFCHLYWG----CTRTGCYGCLAPF-CELNLGDKCLDGVLNNNSYESD-----ILKNY 563
          | ::   :   :| || |: :  | |::|:  |::|          :   :
Db    961 PFSLLFLSRERGKCPSTVCGVCGKPFACKSAL--------EIHYRSHTKERPFVCALCRRG 1013

Qy    564 LATRGLTWKNMLTESLVALQRGVF 587
          | |:    : : :     |:| |
Db   1014 CSTMGNLKQHLLTHRLKELPSQLF 1037
```

```
RESULT 11
P300_HUMAN
ID   P300_HUMAN     STANDARD;      PRT;   2414 AA.
```

```
AC   Q09472;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   E1A-associated protein p300 (EC 2.3.1.48).
GN   EP300 OR P300.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95011587; PubMed=7523245;
RA   Eckner R., Ewen M.E., Newsome D., Gerdes M., Decaprio J.A.,
RA   Lawrence J.B., Livingston D.M.;
RT   "Molecular cloning and functional analysis of the adenovirus E1A-
RT   associated 300-kD protein (p300) reveals a protein with properties of
RT   a transcriptional adaptor.";
RL   Genes Dev. 8:869-884(1994).
RN   [2]
RP   ENZYMATIC ACTIVITY.
RX   MEDLINE=97100994; PubMed=8945521;
RA   Ogryzko V.V., Schiltz R.L., Russanova V., Howard B.H., Nakatani Y.;
RT   "The transcriptional coactivators p300 and CBP are histone
RT   acetyltransferases.";
RL   Cell 87:953-959(1996).
RN   [3]
RP   INTERACTION WITH PCAF.
RX   MEDLINE=96300317; PubMed=8684459;
RA   Yang X.-J., Ogryzko V.V., Nishikawa J.-I., Howard B.H., Nakatani Y.;
RT   "A p300/CBP-associated factor that competes with the adenoviral
RT   oncoprotein E1A.";
RL   Nature 382:319-324(1996).
RN   [4]
RP   INTERACTION WITH HIF1A AND CREBBP.
RX   MEDLINE=97075102; PubMed=8917528;
RA   Arany Z., Huang L.E., Eckner R., Bhattacharya S., Jiang C.,
RA   Goldberg M.A., Bunn H.F., Livingston D.M.;
RT   "An essential role for p300/CBP in the cellular response to hypoxia.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:12969-12973(1996).
RN   [5]
RP   INTERACTIONS WITH NR3C1.
RX   MEDLINE=98250578; PubMed=9590696;
RA   Fryer C.J., Archer T.K.;
RT   "Chromatin remodelling by the glucocorticoid receptor requires the
RT   BRG1 complex.";
RL   Nature 393:88-91(1998).
RN   [6]
RP   INTERACTION WITH NCOA6.
RX   MEDLINE=20283976; PubMed=10823961;
RA   Ko L., Cardona G.R., Chin W.W.;
RT   "Thyroid hormone receptor-binding protein, an LXXLL motif-containing
RT   protein, functions as a general coactivator.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:6212-6217(2000).
RN   [7]
RP   INTERACTION WITH DTX1.
RX   MEDLINE=21576166; PubMed=11564735;
RA   Yamamoto N., Yamamoto S.-I., Inagaki F., Kawaichi M., Fukamizu A.,
RA   Kishi N., Matsuno K., Nakamura K., Weinmaster G., Okano H.,
RA   Nakafuku M.;
RT   "Role of Deltex-1 as a transcriptional regulator downstream of the
RT   Notch receptor.";
RL   J. Biol. Chem. 276:45031-45040(2001).
RN   [8]
RP   PHOSPHORYLATION, AND INTERACTION WITH TCF7L2 AND LEF1.
RX   MEDLINE=22450614; PubMed=12446687;
RA   Hecht A., Stemmler M.P.;
RT   "Identification of a promoter-specific transcriptional activation
RT   domain at the C-terminus of the Wnt-effector protein TCF4.";
RL   J. Biol. Chem. 278:3775-3785(2003).
RN   [9]
RP   VARIANTS BREAST CANCER PRO-827 AND GLY-1013, VARIANT PANCREATIC CANCER
RP   TYR-1650, AND VARIANT COLORECTAL CANCER GLN-2221.
RX   MEDLINE=20164332; PubMed=10700188;
RA   Gayther S.A., Batley S.J., Linger L., Bannister A., Thorpe K.,
RA   Chin S.-F., Daigo Y., Russell P., Wilson A., Sowter H.M.,
RA   Delhanty J.D.A., Ponder B.A.J., Kouzarides T., Caldas C.;
RT   "Mutations truncating the EP300 acetylase in human cancers.";
RL   Nat. Genet. 24:300-303(2000).
RN   [10]
RP   STRUCTURE BY NMR OF 302-418 IN COMPLEX WITH 786-826 OF HIF1A.
RX   MEDLINE=21957254; PubMed=11959990;
RA   Freedman S.J., Sun Z.Y., Poy F., Kung A.L., Livingston D.M.,
RA   Wagner G., Eck M.J.;
RT   "Structural basis for recruitment of CBP/p300 by hypoxia-inducible
RT   factor-1 alpha.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:5367-5372(2002).
CC   -!- FUNCTION: Functions as histone acetyltransferase and regulates
CC       transcription via chromatin remodeling. Acetylates all four core
CC       histones in nucleosomes. Histone acetylation gives an epigenetic
CC       tag for transcriptional activation. Binds to and may be involved
CC       in the transforming capacity of the adenovirus E1A protein.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + histone = CoA + acetyl-histone.
CC   -!- SUBUNIT: Interacts with NR3C1, PCAF, NCOA6, LEF1 and TCF7L2. The
CC       TAZ-type 1 domain interacts with HIF1A. Probably part of a complex
CC       with HIF1A and CREBBP. Interacts with DTX1.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- PTM: Phosphorylated.
CC   -!- DISEASE: Defects in EP300 may play a role in epithelial cancer.
CC   -!- SIMILARITY: Contains 1 bromodomain.
CC   -!- SIMILARITY: Contains 1 ZZ-type zinc finger.
CC   -!- SIMILARITY: Contains 2 TAZ-type zinc fingers.
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/P300ID97.html".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long  as its content is in no  way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U01877; AAA18639.1; -.
DR   PIR; A54277; A54277.
DR   PDB; 1L3E; 11-FEB-03.
DR   TRANSFAC; T01427; -.
DR   Genew; HGNC:3373; EP300.
DR   MIM; 602700; -.
DR   GO; GO:0008022; F:protein C-terminus binding; TAS.
DR   GO; GO:0003713; F:transcription co-activator activity; TAS.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   GO; GO:0007399; P:neurogenesis; TAS.
DR   GO; GO:0007165; P:signal transduction; TAS.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR003101; KIX.
DR   InterPro; IPR000197; TAZ_finger.
DR   InterPro; IPR000433; ZnF_ZZ.
DR   Pfam; PF00439; bromodomain; 1.
DR   Pfam; PF02172; KIX; 1.
DR   Pfam; PF02135; zf-TAZ; 2.
DR   Pfam; PF00569; ZZ; 1.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00551; ZnF_TAZ; 2.
DR   SMART; SM00291; ZnF_ZZ; 1.
DR   PROSITE; PS00633; BROMODOMAIN_1; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
DR   PROSITE; PS50134; ZF_TAZ; 2.
DR   PROSITE; PS01357; ZF_ZZ_1; 1.
DR   PROSITE; PS50135; ZF_ZZ_2; 1.
KW   Transferase; Transcription regulation; Nuclear protein; Bromodomain;
KW   Cell cycle; Zinc-finger; Phosphorylation; Disease mutation.
FT   DOMAIN       11     17       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   ZN_FING     331    417       TAZ-TYPE 1.
```

```
FT   DOMAIN     1067   1139   BROMODOMAIN.
FT   DOMAIN     1572   1818   BINDING REGION FOR E1A ADENOVIRUS.
FT   ZN_FING    1664   1707   ZZ-TYPE.
FT   ZN_FING    1728   1809   TAZ-TYPE 2.
FT   DOMAIN      797    800   POLY-SER.
FT   DOMAIN     1519   1526   POLY-GLU.
FT   DOMAIN     2066   2069   POLY-GLN.
FT   DOMAIN     2190   2195   POLY-GLN.
FT   VARIANT     827    827   L -> P (in breast cancer).
FT                            /FTId=VAR_014428.
FT   VARIANT    1013   1013   E -> G (in breast cancer).
FT                            /FTId=VAR_014429.
FT   VARIANT    1650   1650   S -> Y (in pancreatic cancer).
FT                            /FTId=VAR_014430.
FT   VARIANT    2221   2221   P -> Q (in colorectal cancer).
FT                            /FTId=VAR_014431
SQ   SEQUENCE   2414 AA;  264143 MW;  6BF909E4B9D693 CRC64;

Query Match          3.7%;   Score 132.5;   DB 1;   Length 2414;
Best Local Similarity  21.1%;  Pred. No. 3.4;  Indels 235;  Gaps 31;
Matches 128;  Conservative 75;  Mismatches 235;

Qy     4  PEEGKQSPPPQ-------PWGRLLRLGAEEGEPH--------VLL------RKREWTI   40
Db   300  PNMGQQ-PAPQVQQPGLVTPVAQGMGSGAHTADPEKRKLIQQQLVLLLHAHKCQRRBQAN  358

Qy    41  GRRRGCDLSF---PSNKLVSGDHCRIVVDEKSGQVTLEDTSTGTVINKLKVVKKQTCPL   97
Db   359  GEVRQCNLPHCRTMKNVLNHMTHCQ---SGKSCQVA--HCASSRQIISHWKONCTREHDCPV  413

Qy    98  -----QTGDVIVILVYRKNEPEHNVAYLYESLSEKQGMTQESFE--ANKENVFHGTKDTSG  150
Db   414  CLPLKNAGD-----KRNQQPILTGAPVGLIGNPSSLGVGCQQSAPNLSTVSQIDPSSIERAY  468

Qy   151  AGAGRGADPRVPPSSPATQVCFEE-PQPSTSTSDLFPTASASTEPSPAGRERSSSCGSCSG  209
Db   469  AALGLPYQVNQMPTQPVQVQAKNQNQQPGQSPQGMRPMNMNSA---SPMG------VNG  518

Qy   210  GGGISPKGSGPSVASDEVSSFASALPDR-----KTASFSSLBPQD---QEDLLEPVKKQMR  261
Db   519  GVGVQT----PSLLSD--SMLHSAINSQNPWMSENASVPSLGPMPTAAQPSTTGIRKQWH  572

Qy   262  GDGDLDLNGGLLVAQPRRNAQTVHEDVRAAAGKPD-------KMEETLTCI--ICQDLLH  312
Db   573  EDITQDLRNHL-------VHKLVQAIFPTFPDPAALKDRRMENLVAYARKVEGDMYE  621

Qy   313  DCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNN---LVEAYLIQHPDK  369
Db   622  SANNRAEYYHLLAEKIYKIQKELEBKRRT-------RLQKQNMLPNAAGMVPVSMNPGPNM  675

Qy   370  SRSEEDVQSMD-------ARNKITQDMLQPKVRRSFSDDEGSSEDLLELSDVDSESSDIS  422
Db   676  GQPQPGMTSNGPLPDPSMIRGSVPNQMMPRITPQSGLNQFG------------QMSMA  721

Qy   423  QPYVVCRQCP--EYRRQAAQP--------PHCPAPEGE-------------P  451
Db   722  QPPIVPRQTPLQHGQLAQPGALNPPMGYGPRMQQPSNQGQFLPQTQPPSQGMNVTNIP  781

Qy   452  GAPQALGDAPSTSVSLTTAVQDYVCPL----------QGSHALCTCCFQPMPDRRAERBQD  502
Db   782  LAPSS-GQAPVSQAQMSSS----SCPVNSPIMPFGSQGSHIHCFQLPQF-----ALHQNS  831

Qy   503  PRVAPQO  509
Db   832  PSPVPSR  838

RESULT 12
RN12_MOUSE
ID   RN12_MOUSE   STANDARD;   PRT;   600 AA.
AC   Q9WTV7;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
```

DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RING finger protein 12 (LIM domain interacting RING finger protein)
DE   (RING finger LIM domain-binding protein) (R-LIM).
GN   RNF12 OR RLIM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bach I.;
RT   "Opposing developmental functions of positive and negative
RT   coregulators of LIM homeodomain factors.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   FUNCTION.
RX   MEDLINE=9364422; PubMed=10431247;
RA   Bach I., Rodriguez-Esteban C., Carriere C., Bhushan A., Krones A.,
RA   Rose D.W., Glass C.K., Andersen B., Izpisua-Belmonte J.-C.,
RA   Rosenfeld M.G.;
RT   "RLIM inhibits functional activity of LIM homeodomain transcription
RT   factors via recruitment of the histone deacetylase complex.";
RL   Nat. Genet. 22:394-399(1999).
CC   -!- FUNCTION: Acts as a negative co-regulator for LIM homeodomain
CC   transcription factors. Via the recruitment of the SIN3a/histone
CC   deacetylase corepressor complex.
CC   -!- SUBUNIT: Associates with LIM/homeobox factors.
CC   -!- SIMILARITY: Contains 1 RING-type zinc finger.

DR   EMBL; AF069992; AAD34209.1; -.
DR   MGD; MGI:1342291; Rnf12.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS00518; ZF_RING_1; FALSE_NEG.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Transcription regulation; Zinc-finger.
FT   DOMAIN   415   484   SER-RICH.
FT   ZN_FING   546   587   RING-TYPE.
FT   DOMAIN   447   461   POLY-SER.
SQ   SEQUENCE   600 AA; 66470 MW; E6829953D126E41D CRC64;

```
Query Match          3.7%;   Score 131.5;   DB 1;   Length 600;
Best Local Similarity  23.4%;  Pred. No. 0.7;  Indels 80;  Gaps 11;
Matches 60;  Conservative 29;  Mismatches 87;

Qy   100  GDVIYLVYRKNEPEHNVAYLYESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADP  159
Db   407  GELSYFMYSDSDSEPSA-------SVSSRNVERVES-------RNRGRSGGGNSSG--  449

Qy   160  RVPPSSPATQVCFEEPQPSTSTSDLFPTASASTEPSPAGRERSSSCGSGGGGISPKGSG  219
Db   450  ------------------SSSSSPSPSSGGESSESSSSKMFEGSSEGGS-------SG  482

Qy   220  PS-------VASDEVSSFASALPDRKTASFSSLEPQDQEDLLEPVKKKMRGDGDLDLNG  270
Db   483  PSRKDGRHRAPVTFDE---SGGLPFFSLAQFFLLN-BDDED------------  519

Qy   271  QLLVAQPR-RNAQTVHEDVRAAAGKPDKMEEETLTCIIC--QDLLHDCVSLQPCMHTFCAA  327
Db   520  -----QPRGLTKEQIDNLAMRSFGENDALK----TCSVCITEYTEGDKLRKLPCSHEFHVH  571

Qy   328  CYSGWMERSSLCPTCR  343
Db   572  CIDRWLSENSTCPICR  587
```

```
Qy  152 GAGRGADPRVPPSSPATQVCFEBPQPSTSTSD-----LFPTASASTEPSPAGRERSSSC 206
            |||  ||||  :   ::    :|    :: |:|:|
Db 1033 ----VETPVPPAPATESAIASEQPETRPABTAQPAMEETNANSTETAPKS------- 1080

Qy  207 GSGGGGISPKGSGPSVASDEVSSFASALPDRKTASFSSLEPDQEDLEPVKKMRGDGDL 266
Db 1081 ------------DTATQTENPNSESVPSETTEKVAENPPQENETV--AKNEQEATEPT 1124

Qy  267 DLNGQLIVA-QPRRNAQTVHEDVRAAAGKPD-----KMEETLTCICQDLLHDCVSL- 317
           |||:                                   |::|
Db 1125 PQNGEVAKEDQPTVEANTQTNEATQSEGKTETQTAETKSEPTESVTVSENQPEKTVSQS 1184

Qy  318 ---------QPCMHTFCAACYSGWMERSSLCP-TCRCP------VERICKNHILNN 357
Db 1185 TEDKVVVEKBKAKVET---------EETQKAPQVTSKEPPKQAEPAPEEVPTD---TN 1231

Qy  358 LVEAYLIQH------------PDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEGS 404
              :                :
Db 1232 ABEAQALQQTQPTTVAAAETTSPNSKPAEETQQPSE---KTNAEPVTPVVSENTATQPTE 1288

Qy  405 SEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPHCPAPEGEPG------ 452
           :|                :|   |: |    |||:  ||:
Db 1289 TE---ETAKVEKEKTQ-EVPQVASQESPKQEQPAAKPQAQTKPQAEPARENVLTTKNVGE 1344

Qy  453 -APQALGDAPSTSVSLT 468
Db 1345 PQPQAQPQTQSTAVPTT 1361
```

RESULT 13
IGA4_HAEIN
```
ID   IGA4_HAEIN      STANDARD;      PRT;   1849 AA.
AC   P45386;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN   IGA.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NHTI HK61;
RX   MEDLINE=92234949; PubMed=1373717;
RA   Poulsen K., Reinholdt J., Kilian M.;
RT   "A comparative genetic study of serologically distinct Haemophilus
RT   influenzae type 1 immunoglobulin A1 proteases.";
RL   J. Bacteriol. 174:2913-2921(1992).
-!-  FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC   PRODUCING INTACT FC AND FAB FRAGMNTS.
-!-  CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC   certain Pro-|-Xaa bonds in the hinge region. No small molecule
CC   substrates are known.
-!-  SUBCELLULAR LOCATION: Secreted.
-!-  DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC   SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC   OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC   DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC   SIMILARITY).
-!-  SIMILARITY: Belongs to peptidase family S6.
```
```
DR   EMBL; M87491; AAA24968.1; -.
DR   MEROPS; S06.001;
DR   InterPro; IPR006315; Autotransport.
DR   InterPro; IPR005546; Autotransporter.
DR   InterPro; IPR000710; Peptidase_S6.
DR   InterPro; IPR004899; Pertactin.
DR   Pfam; PF03797; Autotransporter; 1.
DR   Pfam; PF02395; IGA1; 1.
DR   Pfam; PF03212; Pertactin; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   TIGRFAMs; TIGR01414; autotrans_barl; 1.
KW   Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1021       IMMUNOGLOBULIN A1 PROTEASE.
FT   PROPEP     1022   1849       HELPER PEPTIDE (POTENTIAL).
FT   ACT_SITE    299    299       PROBABLE.
SQ   SEQUENCE   1849 AA;  202957 MW;  79A7D018C7150AEA CRC64;
```

```
Query Match            3.7%;  Score 131;  DB 1;  Length 1849;
Best Local Similarity  19.9%;  Pred. No. 3;
Matches   99;  Conservative 67;  Mismatches 185;  Indels 146;  Gaps 22;
```

```
Qy   53 NKLVSGDHCRIVVDEKSG-----QVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVY 107
          ||  ::| |  :|||    ||| |  |:|   |:||:
Db  930 NKSATGNFTLQVAD-KTGEPNHNELTLFDASNA--TRNNLEVT-----LANGSVDRGAW 980

Qy  108 RKNEPEHNVAY-LYESLSEKQGMTQB--------------------SFEANKENVFHGTKDTSGA 151
          :|||  |  :  :||: |:  |
Db  981 KYKLRNVNGRYDLYNPEVEKRNQTVDTTNITTPNDIQADPSAQSNNEBIAR------ 1032
```

RESULT 14
RN27_MOUSE
```
ID   RN27_MOUSE      STANDARD;      PRT;    551 AA.
AC   Q99PJ2; Q99PQ4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   RING finger protein 27 (Glioblastoma-expressed ring finger protein)
DE   (Tripartite motif protein 8).
GN   TRIM8 OR RNF27 OR GERP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20568703; PubMed=11118312;
RA   Vincent S.R., Kwasnicka D.A., Fretier P.;
RT   "A novel RING finger-B box-coiled-coil protein, GERP.";
RL   Biochem. Biophys. Res. Commun. 279:482-486(2000).
RN   [2]
RP   SEQUENCE OF 196-463 FROM N.A.
RX   MEDLINE=21231161; PubMed=11331580;
RA   Reymond A., Meroni G., Fantozzi A., Merla G., Cairo S., Luzi L.,
RA   Riganelli D., Zanaria E., Messali S., Cainarca S., Guffanti A.,
RA   Minucci S., Pelicci P.G., Ballabio A.;
RT   "The tripartite motif family identifies cell compartments.";
RL   EMBO J. 20:2140-2151(2001).
-!-  SUBUNIT: Homodimer (By similarity).
-!-  SUBCELLULAR LOCATION: Nuclear (By similarity).
-!-  DEVELOPMENTAL STAGE: At 10.5 and 12.5 dpc, expressed in the
CC   central nervous system. At 14.5 dpc, expressed in the
CC   and inner neural layer of the retina), in the primitive glomeruli
CC   of the developing kidney, in the villi of the gut and in the
CC   dorsal root ganglia.
-!-  DOMAIN: The coiled coil domain is required for homodimerization
CC   (By similarity).
-!-  SIMILARITY: Contains 1 RING-type zinc finger.
-!-  SIMILARITY: Contains 2 B box-type zinc fingers.
```

```
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AF281047; AAG53088.1; -.
DR   EMBL; AF220035; AAG53489.1; -.
DR   MGD; MGI:1933302; Trim8.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   InterPro; IPR000315; Znf_Bbox.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00643; zf-B_box; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50119; ZF_BBOX; FALSE_NEG.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Nuclear protein; Coiled coil; Zinc-finger; Repeat.
FT   ZN_FING      15     56       RING-TYPE.
FT   ZN_FING      92    132       B BOX-TYPE 1.
FT   ZN_FING     140    182       B BOX-TYPE 2.
FT   DOMAIN      181    249       COILED COIL (POTENTIAL).
FT   DOMAIN      274    295       COILED COIL (POTENTIAL).
FT   CONFLICT    259    259       G -> A (IN REF. 2).
SQ   SEQUENCE   551 AA;  61591 MW;  CECB2D70C8B1E0B6 CRC64;

Query Match             3.6%;  Score 130.5;  DB 1;  Length 551;
Best Local Similarity  23.4%;  Pred. No. 0.72;
Matches   57;  Conservative   21;  Mismatches   79;  Indels   87;  Gaps   11;

Qy      299 EETLTCIICQDLLHDCVSLQPCMHTFCAACY-SGWMERSSL--CPTCRCPVER---ICKN 352
                |:|       |       || ||  :|       |||||    |
Db       10 EEELICPICLHVFVEPVQL-PCKHNFCRGCIGEAWAKDSGLVRCPBCNQAYNQKPGLEKN 68

Qy      353 HILNNLVEAYLIQHPDKS----------------------RSEEDVQSMDARNKITQDML 390
                ||||| ||       |                      :
Db       69 LKLTNIVEKFNALHVEKPPTALHCVFCRRGPPLPAQKVCLRCEAPC----CQSHVQTHLQ 124

Qy      391 QPKVRRSFSDEGSSEDLLELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGE 450
                |:|        || |||   ||:      | ||:   ||    ||::|   ||   |
Db      125 QPSTARGHL----------LVEADDVRAWS---------CPQH--NAYRLYHCRAEQ-- 160

Qy      451 PGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCC------------FQPMEPDRRAE 498
                                                        :|
Db      161 ---------------VAVCQYCCYYSGAHQGHSVCDVEIRRNEIRKMLMKQQERLEE 202

Qy      499 REQD 502
                ||||
Db      203 REQD 206


RESULT 15
ACIN_HUMAN
ID   ACIN_HUMAN      STANDARD;      PRT;  1341 AA.
AC   Q9UKV3; O75158; Q9UG91; Q9UKV1; Q9UKV2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Apoptotic chromatin condensation inducer in the nucleus (Acinus).
GN   ACINUS OR KIAA0670.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1, 2 AND 3), PARTIAL SEQUENCE, FUNCTION,
RP   AND MUTAGENESIS OF ASP-1093.
RX   MEDLINE=99418558; PubMed=10490026;
RA   Sahara S., Aoto M., Eguchi Y., Imamoto N., Yoneda Y., Tsujimoto Y.;
RT   "Acinus is a caspase-3-activated protein required for apoptotic
RT   chromatin condensation.";
```

```
RL   Nature 401:168-173(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 4).
RC   TISSUE=Uterus;
RA   Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RC   TISSUE=Fetal brain;
RA   Li W.B., Gruber C., Jessee J., Polayes D.;
RT   "Full-length cDNA libraries and normalization.";
RL   Submitted (FEB-2003) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 56-1341 FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
-!- FUNCTION: Induces apoptotic chromatin condensation after
    activation by CASP3.
-!- SUBCELLULAR LOCATION: Nuclear.
-!- ALTERNATIVE PRODUCTS:
    Event=Alternative splicing; Named isoforms=4;
    Name=1; Synonyms=L;
      IsoId=Q9UKV3-1; Sequence=Displayed;
    Name=2; Synonyms=S';
      IsoId=Q9UKV3-2; Sequence=VSP_004025, VSP_004028;
    Name=3; Synonyms=S;
      IsoId=Q9UKV3-3; Sequence=VSP_004026, VSP_004029;
    Name=4;
      IsoId=Q9UKV3-4; Sequence=VSP_004027;
-!- TISSUE SPECIFICITY: Ubiquitous.
-!- PTM: Undergoes proteolytic cleavage; the processed form is active,
    contrary to the uncleaved form.
-!- SIMILARITY: Contains 1 SAP domain.
--------------------------------------------------------------------------------
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute.  There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed.  Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
--------------------------------------------------------------------------------
DR   EMBL; AF124726; AAD56724.1; -.
DR   EMBL; AF124727; AAD56725.1; -.
DR   EMBL; AF124728; AAD56726.1; -.
DR   EMBL; AL050382; CAB43681.1; -.
DR   EMBL; BX247975; CAD62309.1; -.
DR   EMBL; AB014570; BAA31645.2; -.
DR   Genew; HGNC:17066; ACINUS.
DR   MIM; 604562; -.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0016887; F:ATPase activity; NAS.
DR   GO; GO:0019899; F:enzyme binding; NAS.
DR   GO; GO:0003676; F:nucleic acid binding; NAS.
DR   GO; GO:0030263; P:apoptotic chromosome condensation; IDA.
DR   GO; GO:0030218; P:erythrocyte differentiation; IEP.
DR   GO; GO:0045657; P:positive regulation of monocyte differentia... .; IEP.
DR   InterPro; IPR003034; SAP.
DR   Pfam; PF02037; SAP; 1.
DR   SMART; SM00513; SAP; 1.
DR   PROSITE; PS50800; SAP; 1.
KW   Apoptosis; Nuclear protein; Alternative splicing.
FT   DOMAIN      -72    106       SAP.
FT   DOMAIN      142    442       GLU-RICH.
FT   DOMAIN      573    676       SER-RICH.
FT   DOMAIN     1114   1131       PRO-RICH.
FT   DOMAIN     1132   1341       ARG/ASP/GLU/LYS-RICH.
```

```
FT   SITE           1093   1093      CLEAVAGE (BY CASPASE-3).
FT   VARSPLIC       1094      1       Missing (in isoform 2).
FT                                    /FTId=VSP_004025.
FT   VARSPLIC        758      1       Missing (in isoform 3).
FT                                    /FTId=VSP_004026.
FT   VARSPLIC       1152      1       Missing (in isoform 4).
FT                                    /FTId=VSP_004027.
FT   VARSPLIC        766    728       GSPKKCEAEEAEPPAATQPQTSETQTSHLPESERIHTV
FT                                    -> MSPADRCRSANTIEPATTSSLALFLLLQRDQSSTRGL
FT                                    P (in isoform 2).
FT                                    /FTId=VSP_004028.
FT   VARSPLIC        766    759       SERIHHTV -> MLSESKEG (in isoform 3).
FT                                    /FTId=VSP_004029.
FT   MUTAGEN        1093   1093      D->A: ABOLISHES CLEAVAGE BY CASP3 AND
FT                                    CHROMATIN CONDENSATION ACTIVITY.
FT   CONFLICT        139    139      Q -> H (IN REF. 4).
SQ   SEQUENCE       1341 AA;  151887 MW;  8FE28668lFB3AB5C CRC64;

Query Match              3.6%;  Score 129;  DB 1;  Length 1341;
Best Local Similarity   20.2%;  Pred. No. 2.7;
Matches   127;  Conservative  78;  Mismatches  239;  Indels  184;  Gaps    30;

Qy      2 ERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRG------CDLSFPSNKL   55
          :  :         ||||||                             |  ||
Db    406 EQTASEEETPPP-----LLTKEASSPPPHPQLHSEE-EIEPMEGPAPPVLIQLSPPN--  456

Qy     56 VSGDHCRIVDEKSGQVT-----LEDTSTSGTVINKLKVVKKQTCPLQTGDVI--YLVYR  108
          ||| ||   |::|:   |                 | :||         | | :   ||||
Db    457 TDADTRELLVSQHTVQLVGGLSPLSSPSDTKAESPAEKVPEESVLPLVQKSTLADYSAQK  516

Qy    109 KNEPEHN------VAYLYESLSEKQGMTQESFE----ANKEN------VFHGTKDTSGA  151
           |  |         :   | | | |      :   ||:|            :|:||
Db    517 DLEPESDRSAQPLPLKIEELALAKGITRECLKQPSLEQKEGRRASHTLLPSHRLKQSADS  576

Qy    152 GAGRGADPRVPFSSPATQVCFEEPQPSTSTSDLFPTAS------ASSTEPSPAG-----  199
          :  :                       :  |           : :|
Db    577 SSSRSSS-----SSSSSRSRSRSPDSSGSRSH-SPLRSKQRDVAQARTHANPRGRPKMGS  631

Qy    200 ------RERSSSCGSGGGGISPKGSGPSVASDEVSSPASALPDRKTASPSSLBPQDQEDL  253
                                 :|:                     : :  ::|
Db    632 RSTSESRSRSRSRSRSASSNSRKSLSPGVSRDSSTSYT---------ETKDPSSGQEV   680

Qy    254 --EPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRA-AAGKPDKMEETLTCIICQDL  310
                      ||            :  :|         | :|
Db    681 ATPVP---------QLQVCEPKERTSTSSSSVQARRLSQPESAEK---------   717

Qy    311 LHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERI---------CKNHILNNLVEA  361
                     :                                     :: |   |
Db    718 -HVTQRLQP-----------ERGS---PKKCEAEEAEPPAATQPQTSETQTSHLPES  759

Qy    362 YLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDI  421
          | :::  :             :: :|::|       |:|:|
Db    760 ERIHHTVEEK--EEVTMDTSENRPENDVPEPMPIA---DQVSNDDRPEGSVEDEEKKES  814

Qy    422 SQPYVVCRQCPEYRRQAAQPP-HCPAPEGEPGAPQALGDAPST-----SVSLTT-----  469
             :                |: :|  |    |     |       |:|:|
Db    815 SLPKSFKRKISVVSATKGVPAGNSDTEGSQGPGRKRRWGASTATTQKKPSISITTBSLKSL  874

Qy    470 ------AVQDYVCPL----------------QGSH----ALCTCCFQPMP--------  493
                    :               |               ::|
Db    875 IPDIKPLAGQEAVVDLHADDSRISEDETERNGDDGTHDKGLKICRTVTQVVPAEBGQENGQ  934

Qy    494 ---DRRAEREQDPRVAPQ-QCAVCLQP  516
             :     |  ::||||
Db    935 REEEEEEKPEAERPVPPQVSVEVALPP  962
```

Search completed: May 7, 2004, 14:46:30
Job time : 26.4372 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 14:40:47 ; Search time 96.0591 Seconds
(without alignments)
2180.991 Million cell updates/sec

Title: US-10-048-046-2
Perfect score: 3585
Sequence: 1 MERPEEGKQSPPQPWGRLL.........VKAHHAMKFNHICEQTRFKN 664

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Query Length	DB	ID	Description
1	3585	100.0	664	4	Q9NRT4	Q9nrt4 homo sapien
2	3504	97.7	652	4	Q96SL3	Q96sl3 homo sapien
3	3500	97.6	652	4	Q96EP1	Q96ep1 homo sapien
4	3162.5	88.2	623	4	Q9NVD5	Q9nvd5 homo sapien
5	2916	81.3	664	11	Q8IOL3	Q8iol3 mus musculu
6	2891.5	80.7	663	11	Q8BJZ9	Q8bjz9 mus musculu
7	1678	46.8	306	6	Q9NT32	Q9nt32 homo sapien
8	896.5	25.0	254	11	Q8BWH4	Q8bwh4 mus musculu
9	357.5	10.0	473	10	Q9SX88	Q9sx88 arabidopsis
10	167.5	4.7	540	13	Q7ZX20	Q7zx20 xenopus lae
11	167	4.7	485	13	Q8O3C1	Q8o3c1 brachydanio
12	160.5	4.5	352	4	Q8BYE7	Q8bye7 homo sapien
13	160.5	4.5	432	4	Q8IUD6	Q8iud6 homo sapien
14	160	4.5	1546	4	Q7Z1L3	Q7z1l3 homo sapien
15	155.5	4.3	733	11	Q9R1A8	Q9r1a8 mus musculu
16	154.5	4.3	731	4	Q8NHY2	Q8nhy2 homo sapien
17	152	4.2	245	11	Q9CQE0	Q9cqe0 mus musculu
18	151.5	4.2	350	4	Q8N3R7	Q8n3r7 homo sapien
19	150	4.2	513	5	Q9CIX4	Q9cix4 schizosacch
20	149	4.2	689	11	Q99PD2	Q99pd2 rattus norv
21	149	4.2	689	10	Q9FNZ7	Q9fnz7 oryza sativ
22	148.5	4.1	209	11	Q9WTR9	Q9wtr9 mus musculu
23	148	4.1	753	11	Q9D4H7	Q9d4h7 mus musculu
24	146.5	4.1	220	4	Q96SJ1	Q96sj1 homo sapien
25	146.5	4.1	277	4	Q96ID9	Q96id9 homo sapien
26	146	4.1	303	4	Q7Z419	Q7z419 homo sapien
27	145	4.0	504	16	Q87PU0	Q87pu0 vibrio para
28	144.5	4.0	2322	10	Q9SV27	Q9sv27 arabidopsis
29	143.5	4.0	638	10	Q8BXP2	Q8bxp2 mus musculu
30	143.5	4.0	700	12	Q7T400	Q7t400 simian herp
31	143.5	4.0	887	11	Q8CFF5	Q8cff5 mus musculu
32	143.5	4.0	1033	11	Q8CGC8	Q8cgc8 mus musculu
33	143.5	4.0	1033	11	Q80Z37	Q80z37 mus musculu
34	143	4.0	1102	5	Q9V5P3	Q9v5p3 drosophila
35	143	4.0	1601	5	Q9V6I9	Q9v6i9 drosophila
36	142	4.0	532	3	Q8WZL0	Q8wzl0 neurospora
37	142	4.0	781	3	Q9C2M7	Q9c2m7 neurospora
38	142	4.0	818	4	Q8WU72	Q8wu72 homo sapien
39	141.5	3.9	424	10	Q9C8W4	Q9c8w4 arabidopsis
40	141.5	3.9	444	10	Q9M621	Q9m621 arabidopsis
41	141.5	3.9	448	10	Q8VZ14	Q8vz14 arabidopsis
42	141.5	3.9	628	13	Q7ZVC1	Q7zvc1 brachydanio
43	141.5	3.9	652	4	Q86Y15	Q86y15 homo sapien
44	141	3.9	1208	4	O75162	O75162 homo sapien
45	140.5	3.9	765	5	Q9UAQ3	Q9uaq3 caenorhabdi

ALIGNMENTS

RESULT 1
Q9NRT4
ID   Q9NRT4   PRELIMINARY;   PRT;   664 AA.
AC   Q9NRT4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cell cycle checkpoint protein CHFR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20388685; PubMed=10935642;
RA   Scolnick D.M., Halazonetis T.D.;
RT   "Chfr defines a mitotic stress checkpoint that delays entry into
RT   metaphase.";
RL   Nature 406:430-435(2000).
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AF170724; AAP91084.1; -.
DR   PDB; 1LGP; 07-AUG-02.
DR   PDB; 1LGQ; 07-AUG-02.
DR   Genew; HGNC:20455; CHFR.
DR   GO; GO:0007093; P:mitotic checkpoint; TAS.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_RING.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 664 AA; 73386 MW; 141A1E7FEFAB36A2 CRC64;

Query Match   100.0%;   Score 3585;   DB 4;   Length 664;

Best Local Similarity 100.0%; Pred. No. 2e-275;
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MERPEBGKQSPPPQPWGRLLRLGAEBGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MERPEBGKQSPPPQPWGRLLRLGAEBGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   60

Qy   61 CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLIVYRKNEPEHNVAYLY  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLIVYRKNEPEHNVAYLY  120

Qy  121 ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEBPQPSTS  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEBPQPSTS  180

Qy  181 TSDLFPTASASTEPSPAGRERSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTA  240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TSDLFPTASASTEPSPAGRERSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTA  240

Qy  241 SFSSLEPQDQEDLEPVKKKWRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE  300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 SFSSLEPQDQEDLEPVKKKWRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE  300

Qy  301 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE  360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE  360

Qy  361 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSEBSD  420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSEBSD  420

Qy  421 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG  480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG  480

Qy  481 SHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELN  540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 SHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELN  540

Qy  541 LGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVL  600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 LGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVL  600

Qy  601 CYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQT  660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 CYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQT  660

Qy  661 RFKN  664
        ||||
Db  661 RFKN  664
```

RESULT 2
Q96SL3

```
ID   Q96SL3         PRELIMINARY;     PRT;     652 AA.
AC   Q96SL3;
DT   01-DEC-2001  (TrEMBLrel. 19, Created)
DT   01-DEC-2001  (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003  (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ14781.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
```

DR   EMBL; AK027687; BAB55297.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_Ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS00089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 652 AA; 72030 MW; 572F2CE6D1743D80 CRC64;

Query Match
Best Local Similarity 97.7%;   Score 3504; DB 4; Length 652;
Matches 652; Conservative 98.2%; Pred. No. 5.2e-269; Mismatches 0; Indels 12; Gaps 1;

```
Qy    1 MERPEBGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MERPEBGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   60

Qy   61 CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLIVYRKNEPEHNVAYLY  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLIVYRKNEPEHNVAYLY  120

Qy  121 ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEBPQPSTS  180
        ||||||||||||||||||||||||||||||
Db  121 ESLSEKQGMTQESF----------DTSGAGAGRGADPRVPPSSPATQVCFEBPQPSTS    168

Qy  181 TSDLFPTASASTEPSPAGRERSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTA  240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  169 TSDLFPTASASTEPSPAGRERSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTA  228

Qy  241 SFSSLEPQDQEDLEPVKKKWRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE  300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  229 SFSSLEPQDQEDLEPVKKKWRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEE  288

Qy  301 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE  360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  289 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE  348

Qy  361 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSD  420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  349 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSD  408

Qy  421 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG  480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  409 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG  468

Qy  481 SHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCBLN  540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  469 SHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCBLN  528

Qy  541 LGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVL  600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  529 LGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDTVL  588

Qy  601 CYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQT  660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  589 CYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQT  648

Qy  661 RFKN  664
        ||||
Db  649 RFKN  652
```

RESULT 3
Q96EP1

```
ID   Q96EP1         PRELIMINARY;     PRT;     652 AA.
AC   Q96EP1;
DT   01-DEC-2001  (TrEMBLrel. 19, Created)
DT   01-DEC-2001  (TrEMBLrel. 19, Last sequence update)
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DB   Hypothetical protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; BC012072; AAH12072.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 652 AA; 72058 MW; 1A5D773CEA679F07 CRC64;

Query Match    97.6%;  Score 3500;  DB 4;  Length 652;
Best Local Similarity  98.0%;  Pred. No. 1.1e-268;
Matches 651;  Conservative 0;  Mismatches 1;  Indels 12;  Gaps 1;

Qy     1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH 60

Qy    61 CRIVDEKSGGVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CRIVDEKSGGVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY 120

Qy   121 ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS 180
          ||||||||||||||         ||||||||||||||||||||||||||||||||
Db   121 ESLSEKQGMTQESF--------DTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS 168

Qy   181 TSDLFPTASASSTEPSPAGRERSSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   169 TSDLFPTASASSTEPSPAGRERSSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTA 228

Qy   241 SFSSLEPDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDRKMEE 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   229 SFSSLEPDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDRKMEE 288

Qy   301 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   289 TLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNLVE 348

Qy   361 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSD 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   349 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSD 408

Qy   421 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   409 ISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQG 468

Qy   481 SHALCTCCFQPMPDRRABREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELN 540
          |||||||||||||||  |||||||||||||||||||||||||||||||||||||||||||
Db   469 SHALCTCCFQPMPDRRVEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELN 528

Qy   541 LGDKCLDGVLNNNSYESDILKNYLATRGLTWKOMLTESLVALQRGVFLLSDYRVTGDTVL 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   529 LGDKCLDGVLNNNSYESDILKNYLATRGLTWKOMLTESLVALQRGVFLLSDYRVTGDTVL 588

Qy   601 CYCCGLRSFRELTYQRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQT 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   589 CYCCGLRSFRELTYQRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICEQT 648
```

```
Qy   661 RFKN 664
         ||||
Db   649 RFKN 652

RESULT 4
Q9NVD5
ID   Q9NVD5            PRELIMINARY;      PRT;   623 AA.
AC   Q9NVD5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ10796.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA   Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA   Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AK001658; BAA91817.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein; Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 623 AA; 69204 MW; 45186D33DAE52711 CRC64;

Query Match    88.2%;  Score 3162.5;  DB 4;  Length 623;
Best Local Similarity  88.5%;  Pred. No. 5.7e-242;
Matches 594;  Conservative 5;  Mismatches 17;  Indels 55;  Gaps 2;

Qy     1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH 60

Qy    61 CRIVDEKSGGVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CRIVDEKSGGVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY 120

Qy   121 ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS 180
          |||||||||||||||
Db   121 ESLSEKQGMTQESFE--------------------------------------------- 135

Qy   181 TSDLFPTASA------SSTEPSPAGRERSSSCGSGGGGISPKGSGPSVASDEVSSFASA 233
                            ::||            |||||||||||||||||||||||||||
Db   136 ---MVPCCVAQAGLKLLGSSDPPTLASQSIVITGSSGGGISPKGSGPSVASDEVSSFASA 192

Qy   234 LPDRKTASFSSLEPDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAG 293
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   193 LPDRKTASFSSLEPDQEDLEPKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAG 252

Qy   294 KPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNH 353
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   253 KPDKMEETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNH 312

Qy   354 ILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSD 413
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   313 ILNNLVEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSD 372
```

Qy   414 VDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQOALGDAPSTSVSLTTAVQD 473
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   373 VDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQOALGDAPPTSVSLTTAVQD 432

Qy   474 YVCPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCL 533
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   433 YVCPLQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCL 492

Qy   534 APFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYR 593
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   493 APFCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYR 552

Qy   594 VTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKP 653
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   553 VTGDTVLCYCCGLRSFRELTYQYQQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKF 612

Qy   654 NHICEQTRFKN 664
          |||||||||||
Db   613 NHICEQTRFKN 623

RESULT 5
Q81OL3
ID   Q81OL3;   PRELIMINARY;   PRT;   664 AA.
AC   Q81OL3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RIKEN cDNA 5730484M20 gene.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Limb;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC049792; AAH49792.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 2.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE   664 AA; 73871 MW; D651BB3E463DBBB6 CRC64;

Query Match   81.3%; Score 2916; DB 11; Length 664;
Best Local Similarity   82.0%; Pred. No. 2.1e-222;
Matches 547; Conservative 41; Mismatches 73; Indels 6; Gaps 4;

Qy   1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   60
          ||||||||||||:||||||||||||||||||:||||||||||||||||||||||||||||
Db   1 MELHGEEQPPPFQEPWGRLLRLGAEEDEPQILLWKREWTIGRRRGCDLSFPSNKLVSGDH   60

Qy   61 CRIVVDEKSGQVTLEDTSTSGGTVINKLKVVKKQTCPLQTGDVIYLYVRKNEPEHNVAYLY 120
          |||:|||:|||||||:||||:||||||||||:||||:|:||||||||:|||||||||||||
Db   61 CKLTVDEISGEVTLEDTSTNGTVINKLQVVKKQTYPLQSGDIIYLYVLVVRKNEPEHNVAYLY 120

Qy   121 ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPSSPATQVCEEPQPSTS 180
          |||||:|||||||||||:|||:||:||||::|:||:||:||||:||||||:||||||||
Db   121 ESLSGKQSLTQDSLEANKENMFHVTKDCS--GPGQGDDPQVPLLSPMAQTCLEEPQPSTS 178

Qy   181 TSDLFPTASASSTEP--SPAGRERSSSCGSGSGGGGISPKGSGPSVASDEVSSFASALPDRK 238
          ||||||||:|||||| |:||:||::|:||||:|:||||:||:|||:||:|||:||:|:|::
Db   179 TSDLLPTASTSSTEPELTSAGQKHSSSSGSPGNTSISPKGRSSLVANGELSSLSPVFQD-K 237

Qy   239 TASFSSLEPQDQEDLEPVKKKNVRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKM 298
          |||:|||||||:|||||:||||:||||||||||||:|||||||:||||||::|:|:||||||

RESULT 6
Q8BJZ9
ID   Q8BJZ9   PRELIMINARY;   PRT;   663 AA.
AC   Q8BJZ9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cell cycle checkpoint protein CHFR.
GN   5730484M20RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Body;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium;
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK077629; BAC36912.1; -.
DR   MGD; MGI:2444898; 5730484M20Rik.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_Ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 2.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE   663 AA; 73855 MW; 42BD653DCB4ADF45 CRC64;

Query Match   80.7%; Score 2891.5; DB 11; Length 663;
Best Local Similarity   81.7%; Pred. No. 1.9e-220;
Matches 545; Conservative 41; Mismatches 74; Indels 7; Gaps 5;

Qy   1 MERPEEGKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   60
          ||||||||||||:||| :|| :||| ::|:||||||| ||||||:||:|||||||| |||

Db   238 EASFSLLESKDHEELBPAKKKMKGDGELDTNLQLLVSGQRGNAQTSSEDVKDASVKPDKM 297
          ||||:||:||::|:|:||::: ::||||:||:|||||| ||:|:||::||:||:|||||||
Qy   299 EETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNL 358
          ||||||||:||||||||||||||||||||||||||||||||||||||||||||||||:||
Db   298 EETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSSLCPTCRCPVERICKNHILNNL 357

Qy   359 VEAYLIQHPDKSRSEEDVQSMDARNKITODMLQPKVRRSFSDEGGSSEDLLELSDVDSES 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db   358 VEAYLIQHPDKSRSEEDVRSMDARNKITODMLQPKVRRSFSDEGGSSEDLLELSDVDSES 417

Qy   419 SDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQAL-GDAPSTSVSLTTAVQDYVCP 477
          ||||||||||||||||||||:|||||||||||||:: :||:||:||:|||:||::||||||
Db   418 SDISQPYIVCRQCPEYRRQAVQSLPCPVPESELGATLALGGEAPSTSASLPTRAPDYMCP 477

Qy   478 LQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFC 537
          |||||||||||:||||||||||||||||||||||||||:||||||||||||||||||||||
Db   478 LQGSHAICTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCFGCLAPFC 537

Qy   538 ELNLGDKCLDGVLNNNYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGD 597
          |||||||||||||||| :||||||||||| |||||||||||| :|||||||:||||:||
Db   538 ELNLGDKCLDGVLNNNYESDILKNYLATRGLTWKSVLITESLIALQRGVFMLSDYRITGN 597

Qy   598 TVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHIC 657
          |||||||||||||||||||||||||||||:|::::||||||||||||||||||||||||||
Db   598 TVLCYCCGLRSFRELTYQYRQNIPASELPVTVTSRPDCYWGRNCRTQVKAHHAMKFNHIC 657

Qy   658 EQTRFKN 664
          |||||||
Db   658 EQTRFKN 664

Matches   306;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy    359  VEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSES  418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  VEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSES   60

Qy    419  SDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPL  478
           ||||||||||||||| : ||||||||||||||||||||:|:|||:|||||||||||||||
Db     61  SDISQPPQVPLLSPMAQTCLEEPQPSTS                                  120

Qy    479  QGSHALCTCCQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGYGCLAPFCE   538
           |||||||||||||||||||:|||:||||||||||||||||||||||||||||||||||
Db    121  QGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGYGCLAPFCE  180

Qy    539  LNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDT  598
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  LNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGDT  240

Qy    599  VLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICE  658
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  VLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHICE  300

Qy    659  QTRFKN  664
           ||||||
Db    301  QTRFKN  306
```

RESULT 8
Q8BWH4

```
ID   Q8BWH4          PRELIMINARY;       PRT;   254 AA.
AC   Q8BWH4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to cell cycle checkpoint protein CHFR (Fragment).
GN   5730484M2ORIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Lung;
RX   MEDLINE=22354683; PubMed=12466851;
RA   the FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK052473; BAC35008.1; -.
DR   MGD; MGI:2444898; 5730484M20Rik.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF00498; FHA; 1.
DR   SMART; SM00240; FHA; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
FT   NON_TER    254    254
SQ   SEQUENCE  254 AA;  27819 MW;  0340571A2BAEB6EF CRC64;
```

Query Match   25.0%;   Score 896.5;   DB 11;   Length 254;
Best Local Similarity   70.0%;   Pred. No. 6.9e-63;
Matches 180; Conservative   23; Mismatches 49; Indels   5; Gaps   3;

```
Qy     1  MRPEEGKQSPPPQPWGRLLRLGAEBGEPHVLLRKREWTIGRRRGCDLSFPSNKLVSGDH   60
          | ::::||::||||||||||||| :| :|||:||||||||||||||||||||
Db     1  MELHGEQPPPQPQBPWGRLLRLGAEBDEPQILLWKREWTIGRRRGCDLSFPSNKLVSGDH   60

Qy    61  CRIVDEKSGVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY   120
          |||:|| ||| ::||:||||||||| : :||| |||:|||:||||||||||||||||
Db    61  CKLTVDBISGEVTLEDTSTNGTVINKLQVVKKQTYPLQSGDIIYLVYRKNEPBHNVAYLY  120

Qy   121  ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS  180
          |||||||||||||||||||||||||| | |||: ||| |  |||||
```

RESULT 7
Q9NT32

```
ID   Q9NT32          PRELIMINARY;       PRT;   306 AA.
AC   Q9NT32;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   DKFZP434N2420.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137561; CAB70812.1; -.
DR   PIR; T46399; T46399.
KW   Hypothetical protein.
FT   NON_TER    1    1
SQ   SEQUENCE  306 AA;  34501 MW;  6B50F04601FB2939 CRC64;
```

Query Match   46.8%;   Score 1678;   DB 4;   Length 306;
Best Local Similarity 100.0%;   Pred. No. 1e-124;

```
Db     1  MELHGEQPPPQPQPWGRLLRLGAEBEDPQILLWKREWTIGRRRGCDLSFPSNKLVSGDH   60

Qy    61  CRIVDEKSGVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVIYLVYRKNEPEHNVAYLY  120
          |:::| ||||| : ||:||||||||| : :||| |||:|||:||||||||||||||||
Db    61  CKLTVDEISGEVTLEDTSTNGTVINKLQVVKKQTYPLHSGDIIYLVYRKNEPBHNVAYLY  120

Qy   121  ESLSEKQGMTQESFEANKENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTS  180
          ||| : :|| ::| ||::||||| ||:||: |: ||| |  |:||||
Db   121  ESLSGKQSLTQDSLEANKENMFHVTKDCS--GPGQGDDPQVPLLSPMAQTCLEEPQPSTS  178

Qy   181  TSDLFPTASASSTEP--SPAGRBRSSSCSGSGGGISPKGSGPSVASDEVSSFASALPDRK  238
          ||| |||:|:||:|   | :||: :: : |:  | | :|:  ::  :::  | :|:|
Db   179  TSDLLPTASTSSTEPELTSAGQKHSSSSGPGNTSISPKGRSSLVANGELSSLSPVFQD-K  237

Qy   239  TASFSSLEPQDQEDLEPVKKKMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKM  298
          ||||:|: | :: |: | ||::|| ||: : |:||:| :| ||:| :: || || ||:|
Db   238  EASFSLLSKDHELEPAKKKMKGDGELDTNLQLLVSGQRGNAQTSSEDVKDASVKFPDKM   297

Qy   299  EETLTCIICQDLLHDCVSLQPCMHTFCAACYSGMERSSLCPTCRCPVERICKNHILNNL  358
          ||||||||||||||||||||||||||||||||||| |||| ||||| |||:|||||||
Db   298  EETLTCIICQDLLHDCVSLQPCMHTFCAACYSGMERSLCPTCRCPVERICKNHILNNL   357

Qy   359  VEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSES  418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   358  VEAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSES  417

Qy   419  SDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQAL-GDAPSTSVSLTTAVQDYVCP  477
          |||||||||||||||||||||||:   :: :|   :: | | |::::::|:|:|: |::|
Db   418  SDISQPYVVCRQCPEYRRQAVQSLPCPVPESELGATTALGGEAPSTSASLPTA-PDYMCP  476

Qy   478  LQGSHALCTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGYGCLAPFC  537
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   477  LQGSHAICTCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGFCCLAPFC  536

Qy   538  ELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVTGD  597
          |||||||||||||||||||||||||||||||||| ::||:|||||||:||:|||||||||
Db   537  ELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKSVLTESLLALQRGVFMLSDYRITGN  596

Qy   598  TVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQVKAHHAMKFNHIC  657
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   597  TVLCYCCGLRSFRELTYQYRQNIPASELPVTVTSRPDCYWGRNCRTQVKAHHAMKFNHIC  656

Qy   658  EQTRFKN  664
          |||||||
Db   657  EQTRFKN  663
```

```
Db   121 ESLSGKQSLTQDSLEANKENMFHVTKDCS--GPGQGDDPQVPLLSPMAQTCLEBPQPSTS 178
Qy   181 TSDLFPTASASSTEP--SPAGRERSSCGSGGGGISPKGSGPSVASDEVSFASALPDRK 238
Db   179 TSDLLPTASTSSTBPELTSAGQKHSSSSGPGNTSISPKGRSSLVANGELSSLSPVFQD-K 237
Qy   239 TASFSSLEPQDQEDLEP 255
Db   238 EASFSLLESKDHEELEP 254

RESULT 9
Q9SX88
ID   Q9SX88           PRELIMINARY;      PRT;    473 AA.
AC   Q9SX88;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   F16N3.15.
GN   F16N3.15.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. columbia;
RA   Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA   Lee J., Liu A., Li J., Kremenetskaia I., Gonzalez A., Altafi H.,
RA   Araujo R., Brooks S., Buehler E., Chao Q., Conway A.B.,
RA   Dunn P., Hansen N., Huizar L., Khan S., Kim C., Palm C., Rowley D.,
RA   Shinn P., Walker M., Davis R.W., Ecker J.R., Federspiel N.A.,
RA   Theologis A.;
RT   "Arabidopsis thaliana chromosome 1 BAC F16N3 sequence.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AC007519; AAD46042.1; -.
DR   PIR; C96516; C96516.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE 473 AA; 53874 MW; D17217493A81C991 CRC64;

Query Match  10.0%;  Score 357.5; DB 10;  Length 473;
Best Local Similarity  23.6%;  Pred. No. 8.4e-20;
Matches 130;  Conservative  65;  Mismatches 201;  Indels 155;  Gaps  20;

Qy   177 PSTSTSDLFPTASASSTEPS----------------------------PAGRERSS 204
         ||:|:: ||  :::||              |:  |:
Db     7 PMTTTAB---TGQSGGSKPSDDDAWAKLVPLDTRFSDIEIRCNDMVICSEIKPSSLEKHE 63
Qy   205 SCGSGGGGISPKGSGPSVASDEVSFASALPDRKTASFSSLEPQDQEDLEPVKKKMRCDG 264
Db    64 WCR------ITKNLGQSSATIHNKSSDAILVDK--------AVVPK------------DG 97
Qy   265 DLD-LNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEETL-----------TCIICQDLLHDC 314
Db    98 AVDIISGSEIVPGPEEQYLQYRFTMPA--PESRTQLLQISIDPEHAKCSICLNIWHDV 155
Qy   315 VSLQPCMHTFCAACYSGWMERSS------LCPTCRCPVERICKNHILNNLVEAYLICHPD 368
Db   156 VTAAPCLHNFCNGCFSEWMRRSEEKHKHVLCPQCRTTVQYVGKNHFLKNIQEEILKVDAA 215
Qy   369 KSRSEEDVQSMDARNKITQDML--QPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPY 425
Db   216 LRRPAEDIAVLDSSASIQSNLIIGSKRKRLNMPAPTHEERDSLRL----------- 261
Qy   426 VVCRQCPEYRRQAAQPPHCPAPEGEPGAPQALGDAPSTSVSLITAVQDYVCPLQGSHALC 485
Db   262 ---------QCPQ---------CVANIG--------------GYRCEHHGAHLQC 284
```

```
Db   486 TCCFQPMPDRRAEREQDPRVAPQQCAVCLQPFCHLYWGCTRTGCYGCLAPFCELN----L 541
             :   : :| : :: :  ::      ::
Db   285 HLCQGMMPFRANLQ-------VPLHCKGCDRPFCGAYWS-SENVTQGVSGPVCVRETFRPI 337
Qy   542 GDKCLDGV----LNNNSYESDILKNYLATRGLTWKNMLTESL-----VALQRGVFLLSDY 592
             :::                     :::
Db   338 SBRTITRIPFITHEMNRHEQDITQRCIAHMEKTVPDVVAEWLRLFNNREIDRSRMPLNHA 397
Qy   593 R-VTGDTVLCYCCGLRSFRELTYQYRQNIPASELPVAVTSRPDCYWGRNCRTQ-VKAHHA 650
             :                 : :::: :  :      :   :::
Db   398 ETITASTHVCNDCYDKLVGFLLYWFRITLPRNHLPADVAAREDCWVGYACRTQHHNEDHA 457
Qy   651 MKFNHICEQTR 661
             |:::|
Db   458 RKRNHVCRPTR 468

RESULT 10
Q7ZX20
ID   Q7ZX20           PRELIMINARY;      PRT;    540 AA.
AC   Q7ZX20;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to ring finger protein (C3HC4 type) 8.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Klein S., Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC046256; AAH46256.1; -.
DR   InterPro; IPR000253; FHA.
DR   InterPro; IPR008984; SMAD_FHA.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00240; FHA; 1.
DR   SMART; SM00184; RING; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
SQ   SEQUENCE 540 AA; 61925 MW; A504B796022A1D94 CRC64;

Query Match   4.7%;  Score 167.5; DB 13;  Length 540;
Best Local Similarity  18.8%;  Pred. No. 0.00012;
Matches 101;  Conservative  84;  Mismatches 216;  Indels 135;  Gaps  .20;

Qy     5 EEGKQSPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRG---------CDLSFPSNKL 55
           :: |:                                      |:|
Db     3 DEG----PGMCW--CLRRCGRNTEDLLLPDGEEVTLGRGLGVTYQLKPTLCPL-----M 50
Qy    56 VSGDHCRIVDEKSGGQVTLEDT-STSGTVINKLKVVKKQTCPLQTGDVIYL-VYRKNEPE 113
           :      : :  :    : :::   : ::||  | :::  | |::|:|
Db    51 ISRTHC-LFKQNTGGEWTVTDNKSLNGVWRNKERLEPHKAYTLSEGALIQLGVPPPNMES 109
Qy   114 HNVAYLY--BSLSBKQGM---TQESFEANKENVPHGTKDTSGAGAGRGADPRVPPSSPA 167
           : ::        : :::             : :: : : :::
Db   110 AEFEYMLVREHLEKLSGSLIRPLPDKTKATRTRKFTSDTDASGNEGPSNFSIPKFYRV 169
Qy   168 TQVCFEPQPSTSTSDLP-PTASASTEPSPAGRERSSSCGSGGGGISP-------- 215
           |  :::  |::  |        | |::  :
Db   170 SRE-DEDSAKSHTTDLYKQPT----VEPTASGTESRLNDSVEAEVVAPTQQQCRSTLQ 223
Qy   216 --------------------KGSGP-------------------- 220
Db   224 LSRVRQTMEIRRLNVQMQEKQMEMQEKLNSPLENQVGANSVLAVQKELRALHNHLSNEQ 283
```

```
Qy  221 -------------SVASDEVSSFASALPDRKTASFSSLEPQ----DQEDLEPVKKKVRGDGD 265
Db  284 EQHMQSVKELKEIFEEEQQSMGS----RKQVEEEHLKEQLAQALQEHTQLMQELNRSKND 339
Qy  266 LDLNGQLLVAQPRRNAQTVHEDVRAAGKP-------DKMEETLTCIICQDLLHDCVSLQ 318
Db  340 FE---QIIEAKNKELQETKEEKEKVFAQKEEVLNHMNDVLDNELQCIICSEHFIEAVTLN 396
Qy  319 PCMHTFCAACYSGWMERSLCPTCRCPVERICKNHIINNLVEAYLIQHPDKSRSEEDVQS 378
Db  397 -CAHSFCSYCIKSWKKRKEECPICRQEIVTETRSLVLDNCIDSMVDKLSPEMKNRRAALI 455
Qy  379 MDARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDISQPYVVCRQCPEY 434
Db  456 LERKEMVQAEESNPVL--VVSDSSFLSDTFYISSSSDSDELGNDFWMRSEEEEY 509
```

RESULT 11
Q803C1
```
ID  Q803C1;            PRELIMINARY;       PRT;    485 AA.
AC  Q803C1;
DT  01-JUN-2003 (TrEMBLrel. 24, Created)
DT  01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Similar to ring finger protein (C3HC4 type) 8.
OS  Brachydanio rerio (Zebrafish) (Danio rerio).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC  Cyprinidae; Danio.
OX  NCBI_TaxID=7955;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Strausberg R.;
RL  Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; BC044545; AAH44545.1; -.
DR  InterPro; IPR000253; FHA.
DR  InterPro; IPR008984; SMAD_FHA.
DR  InterPro; IPR001841; Znf_ring.
DR  Pfam; PF00498; FHA; 1.
DR  Pfam; PF00097; zf-C3HC4; 1.
DR  SMART; SM00240; FHA; 1.
DR  SMART; SM00184; RING; 1.
DR  PROSITE; PS50006; FHA_DOMAIN; 1.
DR  PROSITE; PS00518; ZF_RING_1; 1.
DR  PROSITE; PS50089; ZF_RING_2; 1.
SQ  SEQUENCE  485 AA; 55325 MW;  088A192AB2AE4706 CRC64;
```

Query Match   4.7%;
Best Local Similarity 19.6%;   Pred. No. 0.00011;
Matches 103; Conservative 74; Mismatches 183; Indels 166; Gaps 20;

```
Qy    1 MERPEE-------GKQSPPPQPWGRLLRLGAEEGEPHVLLRKREWTIGRRRGCD--LSFP 51
Db    1 MEKTEEPPSSNNEBDSPAKEKIWCLQRVG-RDSDWLRLFEDSEVSVGRGLNVTHQILSSS 59
Qy   52 SNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGDVI------- 103
Db   60 CPLMISRIHCVFKLNEGRQWTVTDNKSLNGVWVNGKRIPPSTPCILHQSDSVRLGVPLDG 119
Qy  104 ------YLVYRKNPEHNVAYLYESLSEKQGWTQBSFE-ANKENVFHGTKDTSGAGAGRG 156
Db  120 NPVEFDYILVQKNFDDVK-SFLSGNLGKDSGAASLSQKLKNSKRKPDGGDES-------- 170
Qy  157 ADPRVPPSSPATQVCFEEPQPST-STSDLFPTAS--ASSTEPSPAGRER----------- 202
Db  171 ----------EPCFTQHSKSKLYRSSAPDKSRAQPCSGERRETLKLSSRPLE 213
Qy  203 -----------SSSCGSGGGISP------KG--SGPSVASDEVSSFASALPDRKTA 240
Db  214 EDRDKAGSSSSTCSDSSQHLATLHRYNRSLMVLKGRVGDTQXRAABLEQQQTQTPEREK- 272
Qy  241 SFSSLEPQD-QEDLEPVKKKMRG----------------DGDLDLNG 270
```

```
Qy  273 -----EMQDLQTQLEALRGQLRSQOEQALRMETLEKSFCEEBRRLETEKAQQNEVGLKK 327
Qy  271 QL--------------LVAQPRRNAQTVHEDVRAAAGK------PDKME 299
Db  328 QLEEALKEHRKVIEBLKHAWQGFKEVLQAKDKELEVTKEBKEKAKAQKEBVVTQMTEVLE 387
Qy  300 ETLTCIICQDLLHDCVSLQPCMHTFCAACYSGWMERSLCPTCRCPVERICKNHIINNLV 359
Db  388 SELQCSICSELFIEAVTLN-CAHSFCQHCISEWRNRKDKCPMCWQNITSQTRSLVLDNCI 446
Qy  360 EAYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVARSFSDEEGSS 405
Db  447 D--------RMVENL-SADMRERRLVLINERKGERSKSESISSS 481
```

RESULT 12
Q9BYE7
```
ID  Q9BYE7             PRELIMINARY;       PRT;    352 AA.
AC  Q9BYE7;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  HMBLR protein.
GN  HMBLR.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Testis;
RA  Takahashi N., Akasaka T., Nakayama T., Koseki H., Koga H.;
RT  "identification and characterization of hMBLR.";
RL  Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Cervix;
RA  Strausberg R.;
RL  Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
CC  -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR  EMBL; AB047006; BAB40779.1; -.
DR  EMBL; BC010235; AAH10235.1; -.
DR  Genew; HGNC:21156; RNF134.
DR  InterPro; IPR002350; kazal.
DR  InterPro; IPR001841; Znf_ring.
DR  Pfam; PF00097; zf-C3HC4; 1.
DR  SMART; SM00184; RING; 1.
DR  PROSITE; PS00282; KAZAL; 1.
DR  PROSITE; PS00518; ZF_RING_1; 1.
DR  PROSITE; PS50089; ZF_RING_2; 1.
KW  Metal-binding; Zinc; Zinc-finger.
SQ  SEQUENCE  352 AA; 39241 MW;  559F971A86D9A093 CRC64;
```

Query Match   4.5%;
Best Local Similarity 21.4%;   Pred. No. 0.00023;
Matches 62; Conservative 35; Mismatches 128; Indels 65; Gaps 7;

```
Qy  139 ENVFHGTKDTSGAGAGRGADPRVPPSSPATQVCFEEPQPSTSTSDLFPTASASSTEPSPA 198
Db    2 EGVAVVTAGSVGAAKTEGAAALPPP---------PPPPVSPPALTPAPAAGEGPAPL 50
Qy  199 GRERSSSCGSGGGGISPKGSGPSVASDEVSSFASALPDRKTASFSSLEPQDQEDLEPVKK 258
Db   51 SETGAPGC------------SGSRPPELB--------PERSLGRFRGRFEDEDEELEEEE- 90
Qy  259 KMRGDGDLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEET----------LTCI 306
Db   91 ------ELEEBEBEEDMSHFSLRLEGGRQDSDEEERLINLSELTPYILCSI 138
Qy  307 CQDLLHDCVSLQPCMHTFCAACYSGWMERSLCPTCRC-----PVERICKNHIILNNLVE 360
Db  139 CKGYLIDATTITECLHTFCKSCIVRHFYYSNRCPKCNIVHQTQPLYNIRLDRQLQDIVY 198
```

Qy   361 AYLIQHPDKSRSEEDVQSMDARNKITQDMLQPKVRRSFSDEEGSSEDDLE 410
         :::   ::   : :   :::: :::
Db   199 KLVINLEER----EKKQMHDFYKERGLEVPKPKPAVPQPVPSSKGRSKKVLE 244

RESULT 13
Q8IUD6
ID   Q8IUD6;   PRELIMINARY;   PRT;   432 AA.
AC   Q8IUD6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   MGC13061.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21375674; PubMed=11468690;
RA   Jenne D.E., Tinschert S., Reimann H., Lasinger W., Thiel G.,
RA   Hameister H., Kehrer-Sawatzki H.;
RT   "Molecular characterization and gene content of breakpoint boundaries
RT   in NF1 patients with long range deletions.";
RL   Am. J. Hum. Genet. 69:516-527(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Jenne D.E., Tinschert S., Stegmann E., Reimann H., Nuernberg P.,
RA   Horn D., Naumann I., Buske A., Thiel G.;
RT   "A common set of at least 11 functional genes is lost in the majority
RT   of NF1 patients with gross deletions.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Jenne D.E., Dorschner M.O., Stephens K., Hameister H., Tinschert S.,
RA   Kehrer-Sawatzki H.;
RT   "Complete physical map and gene repertoire around the NF1 gene in man
RT   and mice.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ496729; CAD43140.1; -.
DR   Genew; HGNC:21158; RNF135.
DR   InterPro; IPR001870; B302.
DR   InterPro; IPR003877; SPRY_receptor.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00622; SPRY; 1.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   SMART; SM00184; RING; 1.
DR   SMART; SM00449; SPRY; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Hypothetical protein.
SQ   SEQUENCE 432 AA; 47873 MW; A6756ECF81A6A2A6 CRC64;

Query Match   4.5%;   Score 160.5; DB 4;   Length 432;
Best Local Similarity   21.0%;   Pred. No. 0.00031;
Matches   85;   Conservative   45;   Mismatches 158;   Indels 117;   Gaps   16;

Qy   299 EETLTCIICQDLLHDCVSLQPCMHTFCAACYSG-WMERSS---LCPTCR--CPVERICK 351
         :::   : ::: :::: :   : : ::::::
Db   16 EDDLGCIICQGLL-DWPATLPCGHSFCRHCLEALWGARDARRWACPTCRQGAAQQPHLRK 74

Qy   352 NHILNNLVEAY----------------LIQHPDKSRSEEDVQSMDARNKITQ- 387
         ::   ::   :::
Db   75 NTLLQDLADYKYRRAAREIQAGSDPAHCPCPGSSSLSSAAARPRRRPELQRVAVEKSITEV 134

Qy   388 --------------DMLQPKVRRSFSDEEGSSEDLLELSDVDSESSDIS------------- 422
         :   :::
Db   135 AQELTELVEHLVDIVRSLQNQRPLSESGPDNELSIIGKAFSSGVDLSMASPKLVTSDTAA 194

Qy   423 -------------QPYVVCRQCPEYRRQ-----AAQPHCPAPEGEPGAPQALGDA 460
         ::   :: : ::   ::
Db   195 GKIRDIHDLEIQEKLQESVTWKEAPEAQMQGELLEAPSSSSCPLPDQSHPALRRASRF 254

Qy   461 PSTSVSLTTAVQDYVCPLQGS--HALCTCFQPMPDR-RAEREQDPRVAPQQCAVCLQPF 517
              ::   ::   ::::   ::   :::   :: ::
Db   255 AQWAIHPTFNLKSLSCSLEGSKDSRTVTVSHRPQYTRWNCERFSTSQVL---CSQALSSG 311

Qy   518 CHLYWGCTRTGCYGCLAPFCBLNLGDKCLDGVLNNSYESDLIKNYIATRGLTWK----- 572
                  : ::   ::   :   :   ::
Db   312 KGH-YWEVDTRNC-----------------SHWAVGVASWEMSRDQVLGRTMDSCCVEWKGTSQL 357

Qy   573 ---NMLTESLVALQR-----GVFL------LSDYRVTGDTVLCYCC 604
            ::: : :::   :: :   ::
Db   358 SAWHMVKETVLGSDRPGVVGIWLNLEEGKLAPYSVDNQEKLLYEC 402

RESULT 14
Q7Z3L3
ID   Q7Z3L3;   PRELIMINARY;   PRT;   1546 AA.
AC   Q7Z3L3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein DKFZp686H0793 (Fragment).
GN   DKFZP686H0793.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUB=Human retina;
RA   Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,
RA   Fobo G., Han M., Wiemann S.;
RL   Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BX537764; CAD97829.1; -.
KW   Hypothetical protein.
FT   NON_TER   1   1
SQ   SEQUENCE 1546 AA; 170918 MW; 0702441126BAC818 CRC64;

Query Match   4.5%;   Score 160; DB 4;   Length 1546;
Best Local Similarity   21.1%;   Pred. No. 0.0019;
Matches   109;   Conservative   74;   Mismatches 182;   Indels 152;   Gaps   24;

Qy   51 PSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGVINKLKVVKKQTCPLQTGDVIYLVYRKN 110
              :   :   ::: :   :
Db   68 PSHGRVAGFEVQSLQD-AGNQLGMEDTSLSSGML----TQNTNVPILEG------- 111

Qy   111 EPEHNVAVLYESLSEKQGMTQESFEA-NKENVFHGTKDTSGAGAGRGADPRVPPSSPATQ 169
         :: :   :: :::   :   ::   ::
Db   112 ----------VDVAISQGITLPSLESFHPLNIHIGKGKLHATGSKRGKKMTLRP-GPVTQ 160

Qy   170 ------VCFBEQPSTSTSDLFPTASASTEPSPAGRERSSSCCGSGGGGISPKGSGPSVA 223
                ::   :::: ::::
Db   161 EDRCDHLTLKEP------------------FSGPSEEVKEE-----GGKPQMNSEGEIPSLP 200

Qy   224 SDSVSSFASALPDRKT--------------------ASFSSLEPQDQEDLEPV 256
              :::   :   : :   ::: ::: :::
Db   201 SGSQSAKPVSQPRKSTQPDVCASPQEKPLRTLFHQPEEIEDGGLFIPMEEQDNEESEKR 260

Qy   257 KKKMRG-----DG-------DLDLNGQLLVAQPRRNAQTVHEDVRAAAGKPDKMEETL 302
              :::   :   :: ::   :::   :
Db   261 RKKKKGTKRKRDGRGQEGTLAYDLKLDDML----------DRTLEDGAK------QHNL 303

Qy   303 TCIICQDLLHDCVSLQPCMHTFCAA-----------CYSGWMERSSL--------CPTCR-C 344
              ::   :   :: :::::   :::: :
Db   304 TAVNVRNILHEVTINEHVVAMMKAAISTEDMPMFEPKMTRSKLKEVVEKGVVIPTWNIS 363

Qy   345 PVERICKNHIL-NNLVEAVLIQHPDKSRSEEDVQSMDARNKIT--QDMLQPKVRRSFSDE 401
              :::   : :
Db   364 PIKK--ANEIKPQFVDIHL---BEDDSDEEYQPDDBEEDTABESLLESDVESTASSP 418

Qy   402 EGSSEDLL---ELSDVDSESSDISQPYVVCRQCPEYRRQAAQPPHCPAPEGEPGAPQAL 457
              ::: :   ::
Db   419 RGAKKSRLRQSSMTETDEESGILSEAEKV--TAPAIRHISAE----VVPMGPPPPKPK 472

```
Qy  207  FKLDHSVSTNGHRWQIFQDLL-------------------------GTDQDNLDLANVNLMLELLVQKKKQL  254
Db  416  ------------------------------------------------------------SESSDISQPYVVCRQC  431
Db  255  EAESHAAQLQILMEFLKVARRNKREQLEQIQKELSVLEEDIKRVEEMSGLYSPVSEDSTV  314
Qy  432  PEYRRQAAQPPHC--------PAPEGEPGAPQAL------GDAPSTSVSLTTAVQDYVCP  477
          |::  :|:|              |  :|                |||
Db  315  PQF--EAPSPSHSSIIDSTEYSQPPGPSGTSQTKKQPWYNSTLASRRKRLTAHPED----  368
Qy  478  LQGSHALCTCCFQPMPDRRAEREQDPRVAPQ--QCAVCLQPFCHLYWGCTRTGCYGCLAP  535
          |||                 :|   :     |  |||  :|  || |||  :|
Db  369  ------LEQCYFS--TRMSRISDDSRTASQLDEFQECLSKP------TR---YNSVRP  409
Qy  536  FCELNLGDKCLDGVLNNNSYESDILKNYLATRGLTWKNMLTESLVALQRGVFLLSDYRVT  595
          :   |          :|        ||:  : ||     :|   | | |||| | :
Db  410  LATLSYASDLYNGSSIVSSIEFDRDCDYFAIAGVTKKIKVYEYGTVIQDAV----DIHYP  465
Qy  596  GDTVLC----YCCGLRSFRELTYQYRQNIPAS  623
          :   :  |             :||:||
Db  466  ENEMTCNSKISCISWSS------YHKNLLAS  490

Search completed: May  7, 2004, 14:49:01
Job time : 100.059 secs
```

```
Qy  458  GDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQPMPD  494
          | |:  :  :| : |        :|     ||||||
Db  473  QTRDSTFMEKLHAVDEEL---ASSPVCMDSFQPMDD  505

RESULT 15
Q9R1A8
ID   Q9R1A8          PRELIMINARY;      PRT;     733 AA.
AC   Q9R1A8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Constitutive photomorphogenic protein.
GN   COP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=99324249; PubMed=10395541;
RA   Wang H., Kang D., Deng X.W., Wei N.;
RT   "Evidence for functional conservation of a mammalian homologue of the
RT   light-responsive plant protein COP1.";
RL   Curr. Biol. 9:711-714(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RA   Yi C., Deng X.-W.;
RT   "The study of mammalian COP1s.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR   EMBL; AF151110; AAD51094.2; -.
DR   MGD; MGI:1347046; Cop1.
DR   InterPro; IPR001680; WD40.
DR   InterPro; IPR001841; Znf_ring.
DR   Pfam; PF00400; WD40; 6.
DR   Pfam; PF00097; zf-C3HC4; 1.
DR   PRINTS; PR00320; GPROTEINBRPT.
DR   SMART; SM00184; RING; 1.
DR   SMART; SM00320; WD40; 6.
DR   PROSITE; PS00678; WD_REPEATS_1; 1.
DR   PROSITE; PS50082; WD_REPEATS_2; 2
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
DR   PROSITE; PS00518; ZF_RING_1; 1.
DR   PROSITE; PS50089; ZF_RING_2; 1.
KW   Metal-binding; Repeat; WD repeat; Zinc; Zinc-finger.
SQ   SEQUENCE  733 AA;  80441 MW;  894AEA412BACC737 CRC64;

Query Match        4.3%;  Score 155.5; DB 11;  Length 733;
Best Local Similarity  21.3%;  Pred. No. 0.0016;
Matches  122;  Conservative  53;  Mismatches  208;  Indels  189;  Gaps  24;

Qy  153  AGRGADPRVPPSSPATQVCPEBPQPSTSTSDLFPTASASSTEPSPAGRERSSSCGSGGGG  212
          | :|||      |   |  |         ||                    |
Db    7  AGSGGSAGTSPGSSAASSV------TSAS------SSLSSSSFPSPPSVAASAATLVSGGV   52
Qy  213  ISPKGSG------PSVASDEVSSFASALPDRKTASFSSLEPQDQQEDLEPVKKXMRGDGD  265
                 |||      |:  |                               |  |
Db   53  APAAGSGGLGGPGRPVLVAAAVSGGASA--------GGAVSAGQSRLSCAARPSAGVGG  103
Qy  266  LDLNGQLLVAQPRRNAQTV-------HEDVRAAAGKFPDKMEBTLTCIICQDLLHDCVS  316
          :   |           :|||       ||         : :::    :|  ::: :
Db  104  ---SSSSLGSSSRKRPLLVPLCNGLLNSYED----------KSNDFVCPICFDMIERAY-  149
Qy  317  LQPCMHTFCAACYSGWMERSLCPTCRCPVERICKNHILNLVEAYLIQHPDKSRSEBDV  376
          :  |   ||   |::      ||||| ||  :   |:  |:  :     | :::
Db  150  MTKCCGHSFCYKCIHQSLEDNNRCPKCNYVVDNI--DHLYPNFLVNELIL-KQKQRFEEKR  206
Qy  377  QSMD--------ARNKITQDMLQPKVRRSFSDEEGSSEDLLELSDVD------------  415
          :  |||      |::|| |||:
```

```
        GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   May 7, 2004, 14:36:57 ; Search time 14.9546 Seconds
                                   (without alignments)
                                   1379.240 Million cell updates/sec

Title:          US-10-048-046-2_COPY_31_103
Perfect score:  378
Sequence:       1 VLLRKREWTIGRRGCDLSF.........INKLKVVKKQTCPLQTGDVI 73

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       158610 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:     1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    A_Geneseq_29Jan04:*
          1:  geneseqp19808:*
          2:  geneseqp19908:*
          3:  geneseqp20008:*
          4:  geneseqp20018:*
          5:  geneseqp20028:*
          6:  geneseqp2003as:*
          7:  geneseqp2003bs:*
          8:  geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match Length	DB	ID	Description
1	378	100.0 426	4	AAU15856	Aau15856 Human nov
2	378	100.0 426	6	ABU54925	Abu54925 Human nov
3	378	100.0 623	4	AAB93182	Aab93182 Human pro
4	378	100.0 623	5	ABB97233	Abb97233 Novel hum
5	378	100.0 652	4	AAB93168	Aab93168 Human pro
6	378	100.0 664	4	AAB83843	Aab83843 Amino aci
7	378	100.0 664	4	AAB20219	Aab20219 Human Chf
8	304	80.4 92	4	AAU16156	Aau16156 Human nov
9	304	80.4 92	6	ABU55225	Abu55225 Human nov
10	275	72.8 99	4	AAU16157	Aau16157 Human nov
11	275	72.8 99	4	AAU16576	Aau16576 Human nov
12	275	72.8 99	6	ABU55645	Abu55645 Human nov
13	275	72.8 99	6	ABU55226	Abu55226 Human nov
14	107.5	28.4 776	2	AAW73894	Aaw73894 Yeast MEC
15	107.5	28.4 821	2	AAW26664	Aaw26664 Yeast che
16	107.5	28.4 821	5	ABP54942	Abp54942 Saccharom
17	107.5	28.4 821	6	ABR53872	Abr53872 Protein s
18	107.5	28.4 821	7	ABU61612	Abu61612 Yeast SPK
19	99	26.2 699	5	AAB03445	Aab03445 Candida a
20	99	26.2 699	5	ABP73798	Abp73798 Candida a
21	86	22.8 176	5	ABP65604	Abp65604 Bifidobac
22	79	20.9 345	6	ABR59714	Abr59714 Human tra
23	78.5	20.8 175	5	ABP64143	Abp64143 Human ORF
24	78.5	20.8 244	4	AAU16158	Aau16158 Human nov
25	78.5	20.8 244	6	ABU55227	Abu55227 Human nov
26	76.5	20.2 517	5	AAB47789	Aab47789 Protein k
27	72	19.0 199	5	ABG77255	Abg77255 Selected
28	72	19.0 199	5	ABJ11126	Abj11126 Yeast sel
29	72	19.0 350	5	ABJ11042	Abj11042 Yeast sel
30	71	18.8 409	6	ABU36268	Abu36268 Protein e
31	70.5	18.7 874	6	ABU17261	Abu17261 Protein e
32	70.5	18.7 877	6	ADA36875	Ada36875 Acinetoba
33	68	18.0 399	4	AAB92565	Aab92565 Human pro
34	68	18.0 754	3	AAY51669	Aay51669 Human nib
35	68	18.0 754	3	AAY32373	Aay32373 Nibrin, e
36	66.5	17.6 288	4	AAG89798	Aag89798 C glutami
37	66	17.5 131	3	AAG01640	Aag01640 Human.sec
38	66	17.5 288	2	AAW95501	Aaw95501 B. subtil
39	66	17.5 992	2	AAW69742	Aaw69742 SAPAP2 pr
40	66	17.5 992	2	AAW69740	Aaw69740 SAPAP1 pr
41	65	17.2 945	4	ABB63213	Abb63213 Drosophil
42	63	16.7 138	5	ABB52816	Abb52816 Escherich
43	62.5	16.5 527	5	ABU05697	Abu05697 M. tuberc
44	62.5	16.5 528	4	AAG81105	Aag81105 Mycobacte
45	62	16.4 1724	4	ABB11806	Abb11806 Human sec

ALIGNMENTS

```
RESULT 1
AAU15856
ID   AAU15856 standard; protein; 426 AA.
XX
AC   AAU15856;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 809.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmological; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0209963P.
PR   26-JUL-2000; 2000US-0209964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
```

```
PR   14-AUG-2000;  2000US-0225211P.        PR   08-NOV-2000;  2000US-0246523P.
PR   14-AUG-2000;  2000US-0225214P.        PR   08-NOV-2000;  2000US-0246524P.
PR   14-AUG-2000;  2000US-0225266P.        PR   08-NOV-2000;  2000US-0246525P.
PR   14-AUG-2000;  2000US-0252267P.        PR   08-NOV-2000;  2000US-0246526P.
PR   14-AUG-2000;  2000US-0225268P.        PR   08-NOV-2000;  2000US-0246527P.
PR   14-AUG-2000;  2000US-0225270P.        PR   08-NOV-2000;  2000US-0246528P.
PR   14-AUG-2000;  2000US-0254477P.        PR   08-NOV-2000;  2000US-0245532P.
PR   14-AUG-2000;  2000US-0225757P.        PR   08-NOV-2000;  2000US-0246609P.
PR   14-AUG-2000;  2000US-0225758P.        PR   08-NOV-2000;  2000US-0246610P.
PR   14-AUG-2000;  2000US-0225759P.        PR   08-NOV-2000;  2000US-0246611P.
PR   18-AUG-2000;  2000US-0226279P.        PR   08-NOV-2000;  2000US-0246613P.
PR   22-AUG-2000;  2000US-0226681P.        PR   17-NOV-2000;  2000US-0249207P.
PR   22-AUG-2000;  2000US-0226868P.        PR   17-NOV-2000;  2000US-0249208P.
PR   22-AUG-2000;  2000US-0227182P.        PR   17-NOV-2000;  2000US-0249209P.
PR   23-AUG-2000;  2000US-0227009P.        PR   17-NOV-2000;  2000US-0249210P.
PR   30-AUG-2000;  2000US-0228924P.        PR   17-NOV-2000;  2000US-0249211P.
PR   01-SEP-2000;  2000US-0229287P.        PR   17-NOV-2000;  2000US-0249212P.
PR   01-SEP-2000;  2000US-0229343P.        PR   17-NOV-2000;  2000US-0249213P.
PR   01-SEP-2000;  2000US-0229344P.        PR   17-NOV-2000;  2000US-0249214P.
PR   05-SEP-2000;  2000US-0229345P.        PR   17-NOV-2000;  2000US-0249215P.
PR   05-SEP-2000;  2000US-0229509P.        PR   17-NOV-2000;  2000US-0249216P.
PR   05-SEP-2000;  2000US-0229513P.        PR   17-NOV-2000;  2000US-0249217P.
PR   06-SEP-2000;  2000US-0304377P.        PR   17-NOV-2000;  2000US-0249218P.
PR   06-SEP-2000;  2000US-0304438P.        PR   17-NOV-2000;  2000US-0249244P.
PR   08-SEP-2000;  2000US-0231242P.        PR   17-NOV-2000;  2000US-0249245P.
PR   08-SEP-2000;  2000US-0231243P.        PR   17-NOV-2000;  2000US-0249264P.
PR   08-SEP-2000;  2000US-0231244P.        PR   17-NOV-2000;  2000US-0249265P.
PR   08-SEP-2000;  2000US-0231413P.        PR   17-NOV-2000;  2000US-0249297P.
PR   08-SEP-2000;  2000US-0231414P.        PR   17-NOV-2000;  2000US-0249299P.
PR   08-SEP-2000;  2000US-0232080P.        PR   17-NOV-2000;  2000US-0249300P.
PR   08-SEP-2000;  2000US-0232081P.        PR   01-DEC-2000;  2000US-0250160P.
PR   12-SEP-2000;  2000US-0231968P.        PR   01-DEC-2000;  2000US-0250391P.
PR   14-SEP-2000;  2000US-0232397P.        PR   05-DEC-2000;  2000US-0251030P.
PR   14-SEP-2000;  2000US-0232398P.        PR   05-DEC-2000;  2000US-0251988P.
PR   14-SEP-2000;  2000US-0232399P.        PR   05-DEC-2000;  2000US-0256719P.
PR   14-SEP-2000;  2000US-0232400P.        PR   06-DEC-2000;  2000US-0251479P.
PR   14-SEP-2000;  2000US-0232401P.        PR   08-DEC-2000;  2000US-0251856P.
PR   14-SEP-2000;  2000US-0330064P.        PR   08-DEC-2000;  2000US-0251858P.
PR   14-SEP-2000;  2000US-0330065P.        PR   08-DEC-2000;  2000US-0251869P.
PR   21-SEP-2000;  2000US-0234223P.        PR   08-DEC-2000;  2000US-0251989P.
PR   21-SEP-2000;  2000US-0234274P.        PR   08-DEC-2000;  2000US-0251990P.
PR   25-SEP-2000;  2000US-0349997P.        PR   11-DEC-2000;  2000US-0254097P.
PR   25-SEP-2000;  2000US-0349998P.        PR   05-JAN-2001;  2001US-0259678P.
PR   26-SEP-2000;  2000US-0355484P.        XX
PR   27-SEP-2000;  2000US-0355834P.        PA   (HUMA-) HUMAN GENOME SCI INC.
PR   27-SEP-2000;  2000US-0235836P.        XX
PR   29-SEP-2000;  2000US-0236327P.        PI   Rosen CA,   Barash SC,   Ruben SM;
PR   29-SEP-2000;  2000US-0236367P.        XX
PR   29-SEP-2000;  2000US-0236368P.        DR   WPI; 2001-488783/53.
PR   29-SEP-2000;  2000US-0236369P.        DR   N-PSDB; AAS25843.
PR   29-SEP-2000;  2000US-0236370P.        XX
PR   02-OCT-2000;  2000US-0236702P.        PT   New nucleic acid molecules encoding 461 human secreted proteins for
PR   02-OCT-2000;  2000US-0240960P.        PT   diagnosing, preventing, treating or ameliorating medical conditions and
PR   02-OCT-2000;  2000US-0241221P.        PT   used as food additives or preservatives.
PR   02-OCT-2000;  2000US-0237037P.        XX
PR   02-OCT-2000;  2000US-0237038P.        PS   Claim 11; SEQ ID NO 809; 980pp; English.
PR   02-OCT-2000;  2000US-0237039P.        XX
PR   02-OCT-2000;  2000US-0237040P.        CC   The invention relates to isolated nucleic acid molecules and their
PR   13-OCT-2000;  2000US-0239935P.        CC   encoded secreted proteins. The nucleic acids and proteins are used to
PR   13-OCT-2000;  2000US-0239937P.        CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
PR   20-OCT-2000;  2000US-0417785P.        CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
PR   20-OCT-2000;  2000US-0241786P.        CC   in diagnosing a pathological condition or susceptibility to a
PR   20-OCT-2000;  2000US-0241787P.        CC   pathological condition. Antibodies to the proteins can also be used in
PR   20-OCT-2000;  2000US-0241808P.        CC   alleviating symptoms associated with the disorders and in diagnostic
PR   20-OCT-2000;  2000US-0241809P.        CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
PR   20-OCT-2000;  2000US-0241826P.        CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
PR   01-NOV-2000;  2000US-0244617P.        CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
PR   08-NOV-2000;  2000US-0246474P.        CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
PR   08-NOV-2000;  2000US-0246475P.        CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
PR   08-NOV-2000;  2000US-0246476P.        CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
PR   08-NOV-2000;  2000US-0246477P.        CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
PR   08-NOV-2000;  2000US-0246478P.        CC   and many other disorders listed in the specification. The polypeptides
                                           CC   can also be used to aid wound healing and epithelial cell proliferation,
```

```
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed

Query Match     100.0%;  Score 378;  DB 4;  Length 426;
Best Local Similarity  100.0%;   Pred. No. 1.5e-40;
Matches  73;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 VLLRKREWTIGRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33 VLLRKREWTIGRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 92

Qy   61 KKQTCPLQTGDVI 73
        |||||||||||||
Db   93 KKQTCPLQTGDVI 105

RESULT 2
ABU54925
ID   ABU54925 standard; protein; 426 AA.
XX
AC   ABU54925;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polypeptide #12.
XX
KW   Human; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0209963P.
PR   26-JUL-2000; 2000US-0209964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226647P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.

PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   N-PSDB; ABX73184.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 11; SEQ ID NO 809; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC   ABU55748 represent human novel polypeptides of the invention
XX
SQ   Sequence 426 AA;

Query Match     100.0%;  Score 378;  DB 6;  Length 426;
Best Local Similarity  100.0%;   Pred. No. 1.5e-40;
Matches  73;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 VLLRKREWTIGRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33 VLLRKREWTIGRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 92

Qy   61 KKQTCPLQTGDVI 73
        |||||||||||||
Db   93 KKQTCPLQTGDVI 105
```

RESULT 3
AAB93182
ID   AAB93182 standard; protein; 623 AA.
XX
AC   AAB93182;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human protein sequence SEQ NO:12128.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 12128; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are also useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH11628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 623 AA;

Query Match   100.0%; Score 378; DB 4; Length 623;
Best Local Similarity   100.0%; Pred. No. 2.5e-40;
Matches   73; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 90

Qy   61 KKQTCPLQTGDVI 73
              |||||||||||||
Db   91 KKQTCPLQTGDVI 103

RESULT 4
ABB97233
ID   ABB97233 standard; protein; 623 AA.
XX
AC   ABB97233;
XX
DT   27-JUN-2002 (first entry)
XX
DE   Novel human protein SEQ ID NO: 501.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag.
XX
OS   Homo sapiens.
XX
PN   WO20222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US026015.
XX
PR   11-SEP-2000; 2000US-00659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   N-PSDB; ABN32419.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis.
XX
PS   Example 2; SEQ ID NO 501; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC   and cancer, to screen for drug, to treat inflammatory conditions e.g.
CC   rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a protein of the invention
XX
SQ   Sequence 623 AA;

Query Match   100.0%; Score 378; DB 5; Length 623;
Best Local Similarity   100.0%; Pred. No. 2.5e-40;
Matches   73; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 90

Qy   61 KKQTCPLQTGDVI 73
              |||||||||||||
Db   91 KKQTCPLQTGDVI 103

RESULT 5
AAB93168
ID   AAB93168 standard; protein; 652 AA.
XX

Left column

```
AC   AAB3168;
XX
DT   26-JUN-2001 (first entry)
XX
DE   Human protein sequence SEQ ID NO:12100.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000;   2000JP-00118776.
PR   02-MAY-2000;   2000JP-00183767.
PR   09-JUN-2000;   2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 12100; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 652 AA;

Query Match        100.0%;  Score 378;  DB 4;  Length 652;
Best Local Similarity  100.0%;  Pred. No. 2.6e-40;
Matches   73;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps    0;

Qy     1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    31 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 90

Qy    61 KKQTCPLQTGDVI 73
         |||||||||||||
Db    91 KKQTCPLQTGDVI 103
```

Right column

```
RESULT 6
AAB83843
ID   AAB83843 standard; protein; 664 AA.
XX
AC   AAB83843;
XX
DT   22-AUG-2001 (first entry)
XX
DE   Amino acid sequence of a human ring finger protein designated FHAR1.
XX
KW   FHAR1; RING finger protein; cancer; vaccine.
OS   Homo sapiens.
XX
PN   WO200142430-A1.
XX
PD   14-JUN-2001.
XX
PF   07-DEC-2000; 2000WO-US033094.
XX
PR   08-DEC-1999;   99US-00456876.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Zhou B, Zhu Y, Chaturvedi P, Hurle MR, Li X;
XX
DR   WPI; 2001-381663/40.
DR   N-PSDB; AAF89709.
XX
PT   New FHAR1 polypeptide, a member of the RING finger protein family for
PT   diagnosing and treating cancer, and for use in anti-cancer vaccines.
XX
PS   Claim 1; Page 19; 28pp; English.
XX
CC   The present sequence represents a FHAR1 polypeptide, which is a member of
CC   the RING finger protein family. FHAR1 is useful in the treatment of
CC   cancer, and as a vaccine for inducing an immunological response in a
CC   mammal. FHAR1 polynucleotides may also be used as a diagnostic reagent
CC   through detection of mutations in the associated gene, and for chromosome
CC   localization studies, and tissue expression studies. FHAR1 antibodies are
CC   useful to isolate and to identify clones expressing the polypeptides, or
CC   to purify the polypeptides by affinity chromatography and to treat cancer
XX
SQ   Sequence 664 AA;

Query Match        100.0%;  Score 378;  DB 4;  Length 664;
Best Local Similarity  100.0%;  Pred. No. 2.7e-40;
Matches   73;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps    0;

Qy     1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    31 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 90

Qy    61 KKQTCPLQTGDVI 73
         |||||||||||||
Db    91 KKQTCPLQTGDVI 103

RESULT 7
AAB20219
ID   AAB20219 standard; protein; 664 AA.
XX
AC   AAB20219;
XX
DT   14-MAY-2001 (first entry)
XX
DE   Human Chfr (checkpoint with FHA and ring finger) protein.
XX
KW   Checkpoint with forkhead associated domain and ring finger; Chfr; human;
KW   mitosis; cell cycle; tumour; diagnosis; antitumour; drug screening;
KW   ubiquitin-protein ligase.
XX
```

```
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          31..103
FT                   /label= Forkhead-associated_domain
FT   Domain          303..346
FT                   /label= Ring_finger-domain
FT   Region          476..641
FT                   /note= "cysteine-rich region"
FT   Misc-difference 580
FT                   /note= "Met in U2OS cells"
XX
PN   WO20010150-A2.
XX
PD   08-FEB-2001.
XX
PF   14-JUN-2000; 2000WO-US016391.
XX
PR   29-JUL-1999; 99US-0146194P.
XX
PA   (WIST-) WISTAR INST ANATOMY & BIOLOGY.
XX
PI   Halazonetis T, Scolnick D;
XX
DR   WPI; 2001-182927/18.
DR   N-PSDB; AAF30352.
XX
PT   Novel nucleic acid sequence of mitotic checkpoint gene encoding a
PT   checkpoint with forkhead-associated domain and ring finger protein, for
PT   diagnosing tumorigenic cells and in screening for anticancer drugs.
XX
PS   Claim 8(a); Fig 4A-C; 85pp; English.
XX
CC   The present sequence is that of human mitotic checkpoint protein Chfr,
CC   having a forkhead associated domain (FHA) and a ring finger domain. The
CC   protein is required for regulation of the transition of cells from
CC   prophase to metaphase during mitosis. It has ubiquitin-protein ligase
CC   activity. The Chfr checkpoint was evident in primary human cells, but was
CC   inactivated in 4 of 8 human cancer cell lines. In U2OS cells, a mutation
CC   was identified that caused a Val to Met amino acid substitution in the
CC   highly conserved C-terminal Cys-rich region of the Chfr protein. In the
CC   absence of the Chfr checkpoint, cells subjected to mitotic stress
CC   condensed their chromosomes despite failing to separate their
CC   chromosomes. Chfr may monitor centrosome separation. Inactivation of the
CC   chfr gene (see AAF30352) in human cancer is theorised to underlie the
CC   increased sensitivity of cancer cells to antimitotic drugs. Polypeptides
CC   comprising the present sequence, or sequences comprising at least amino
CC   acids 11-103, 303-346 and/or 476-641 of this sequence, are claimed.
CC   Claimed methods of determining the tumourigenic potential of a cell
CC   comprise examining the cell for the presence of Chfr expression or for
CC   the presence of Chfr-mediated ubiquitin-protein ligase activity (in both
CC   cases, absence of expression indicating predisposition to tumourigenesis
CC   upon exposure to mitotic stress). A diagnostic kit for detecting the
CC   tumourigenic potential of cell cells comprises a ligand that
CC   binds to Chfr, such as an antibody or its fragment. Inhibitors of Chfr
CC   are identified by monitoring their effect on Chfr expression, and are
CC   used to retard the growth of cancer cells
XX
SQ   Sequence 664 AA;

Query Match             100.0%;  Score 378;  DB 4;  Length 664;
Best Local Similarity   100.0%;  Pred. No. 2.7e-40;
Matches   73;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1  VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     31  VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 90

Qy     61  KKQTCPLQTGDVI 73
           |||||||||||||
Db     91  KKQTCPLQTGDVI 103
```

```
RESULT 8
AAU16156
ID   AAU16156 standard; protein; 92 AA.
XX
AC   AAU16156;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 1109.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmological; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2001; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214486P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216680P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
```

```
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0246474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.

PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251198P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   N-PSDB; AAS26143.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 11; SEQ ID NO 1109; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed

Query Match             80.4%;  Score 304;  DB 4;  Length 92;
Best Local Similarity   98.3%;
Matches   59;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy          1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVV 60
              ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db         33 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTXTSGTVINKLKVV 92
```

RESULT 9
ABU55225

```
ID   ABU55225   standard; protein; 92 AA.
XX
AC   ABU55225;
XX
DT   18-MAR-2003   (first entry)
```

```
XX
DE   Human novel polypeptide #312.
XX
KW   Human; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0209963P.
PR   14-AUG-2000; 2000US-0209964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0231413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.

XX
PI   Rosen CA, Ruben SM, Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   N-PSDB; ABX73484.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, cardiovascular or
PT   renal disorders.
XX
PS   Claim 11; SEQ ID NO 1109; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Epstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC   ABU55748 represent human novel polypeptides of the invention
XX
SQ   Sequence 92 AA;

Query Match             80.4%;  Score 304;  DB 6;  Length 92;
Best Local Similarity   98.3%;  Pred. No. 1.1e-31;
Matches   59;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVDEKSGQVTLEDTSTSGTVINKLKVV 60
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         33 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVDEKSGQVTLEDTXTSGTVINKLKVV 92

RESULT 10
AAU16157
ID   AAU16157 standard; protein; 99 AA.
XX
AC   AAU16157;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human novel secreted protein, Seq ID 1110.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmological; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
```

PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226968P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0210437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0230063P.
PR 14-SEP-2000; 2000US-0230064P.
PR 14-SEP-2000; 2000US-0230065P.
PR 21-SEP-2000; 2000US-0242223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.

PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251889P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI, 2001-488783/53.
DR N-PSDB; AAS26144.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
PS Claim 11; SEQ ID NO 1110; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their

```
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed

Query Match        72.8%;  Score 275;  DB 4;  Length 99;
Best Local Similarity  90.3%;  Pred. No. 7.3e-28;
Matches   56;  Conservative   1;  Mismatches   3;  Indels   2;  Gaps   1;

Qy       12  RRRGCDLSPPSNKLVSGDHCRIVDEKSGQVTLEDTSTSGTVINKLKVKKQTCPLQTGD  71
             | :| | |  ||||||||||||||||||||||||||||||| | |||||||||||||||
Db        9  RDKGPD--FPXNKLVSGDHCRIVDEKSGQVTLEDTSTSGTVINKLKVKKQCPLQTGD   66

Qy       72  VI  73
             ||
Db       67  VI  68

RESULT 11
AAU16576
ID   AAU16576 standard; protein; 99 AA.
XX
AC   AAU16576;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human novel secreted protein, Seq ID 1529.
XX
KW   Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW   cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW   antibacterial; virucide; fungicide; opthalmalogical; vulnerary;
KW   secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW   cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW   cerebral ischaemia; angiogenesis; nervous system disorder;
KW   Alzheimer's disease; infection; ocular disorder; corneal infection;
KW   wound healing; epithelial cell proliferation; skin ageing; food additive;
KW   preservative; antiproliferative.
XX
OS   Homo sapiens.
XX
PN   WO200155322-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001341.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180528P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
```

PR	07-JUN-2000;	2000US-0209467P.
PR	28-JUN-2000;	2000US-0214886P.
PR	30-JUN-2000;	2000US-0215135P.
PR	07-JUL-2000;	2000US-0216647P.
PR	07-JUL-2000;	2000US-0216880P.
PR	11-JUL-2000;	2000US-0217487P.
PR	11-JUL-2000;	2000US-0217496P.
PR	14-JUL-2000;	2000US-0218290P.
PR	26-JUL-2000;	2000US-0220963P.
PR	26-JUL-2000;	2000US-0220964P.
PR	14-AUG-2000;	2000US-0224518P.
PR	14-AUG-2000;	2000US-0224519P.
PR	14-AUG-2000;	2000US-0225213P.
PR	14-AUG-2000;	2000US-0225214P.
PR	14-AUG-2000;	2000US-0225266P.
PR	14-AUG-2000;	2000US-0225267P.
PR	14-AUG-2000;	2000US-0225268P.
PR	14-AUG-2000;	2000US-0225270P.
PR	14-AUG-2000;	2000US-0225447P.
PR	14-AUG-2000;	2000US-0225757P.
PR	14-AUG-2000;	2000US-0225758P.
PR	14-AUG-2000;	2000US-0225759P.
PR	18-AUG-2000;	2000US-0226279P.
PR	22-AUG-2000;	2000US-0226681P.
PR	22-AUG-2000;	2000US-0226868P.
PR	22-AUG-2000;	2000US-0227182P.
PR	23-AUG-2000;	2000US-0227009P.
PR	30-AUG-2000;	2000US-0228924P.
PR	01-SEP-2000;	2000US-0229287P.
PR	01-SEP-2000;	2000US-0229343P.
PR	01-SEP-2000;	2000US-0229344P.
PR	01-SEP-2000;	2000US-0229345P.
PR	05-SEP-2000;	2000US-0229509P.
PR	05-SEP-2000;	2000US-0229513P.
PR	06-SEP-2000;	2000US-0230437P.
PR	06-SEP-2000;	2000US-0230438P.
PR	08-SEP-2000;	2000US-0231242P.
PR	08-SEP-2000;	2000US-0231243P.
PR	08-SEP-2000;	2000US-0231244P.
PR	08-SEP-2000;	2000US-0231413P.
PR	08-SEP-2000;	2000US-0231414P.
PR	08-SEP-2000;	2000US-0232080P.
PR	08-SEP-2000;	2000US-0232081P.
PR	12-SEP-2000;	2000US-0231968P.
PR	14-SEP-2000;	2000US-0232397P.
PR	14-SEP-2000;	2000US-0232398P.
PR	14-SEP-2000;	2000US-0232399P.
PR	14-SEP-2000;	2000US-0232400P.
PR	14-SEP-2000;	2000US-0232401P.
PR	14-SEP-2000;	2000US-0233063P.
PR	14-SEP-2000;	2000US-0233064P.
PR	14-SEP-2000;	2000US-0233065P.
PR	21-SEP-2000;	2000US-0234223P.
PR	21-SEP-2000;	2000US-0234274P.
PR	25-SEP-2000;	2000US-0234997P.
PR	25-SEP-2000;	2000US-0234998P.
PR	26-SEP-2000;	2000US-0235448P.
PR	27-SEP-2000;	2000US-0235834P.
PR	27-SEP-2000;	2000US-0235836P.
PR	29-SEP-2000;	2000US-0236327P.
PR	29-SEP-2000;	2000US-0236367P.
PR	29-SEP-2000;	2000US-0236368P.
PR	29-SEP-2000;	2000US-0236369P.
PR	29-SEP-2000;	2000US-0236370P.
PR	02-OCT-2000;	2000US-0236802P.
PR	02-OCT-2000;	2000US-0237037P.
PR	02-OCT-2000;	2000US-0237038P.
PR	02-OCT-2000;	2000US-0237039P.
PR	02-OCT-2000;	2000US-0237040P.
PR	13-OCT-2000;	2000US-0239935P.
PR	13-OCT-2000;	2000US-0239937P.
PR	20-OCT-2000;	2000US-0240960P.
PR	20-OCT-2000;	2000US-0241221P.

```
CC   in diagnosing a pathological condition or susceptibility to a
CC   pathological condition. Antibodies to the proteins can also be used in
CC   alleviating symptoms associated with the disorders and in diagnostic
CC   immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC   (ELISA). Disorders which are diagnosed or treated include autoimmune
CC   diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC   neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC   arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC   nervous system disorders e.g. Alzheimer's disease, infections caused by
CC   bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC   and many other disorders listed in the specification. The polypeptides
CC   can also be used to aid wound healing and epithelial cell proliferation,
CC   to prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, for supporting cell culture of primary tissues, to
CC   regenerate tissues and in chemotaxis. The polypeptides can also be used
CC   as a food additive or preservative to increase or decrease storage
CC   capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC   minerals, cofactors and other nutritional components. The present
CC   sequence represents a novel secreted protein of the invention. Note: The
CC   sequence data for this patent did not form part of the printed

Query Match    72.8%;  Score 275;  DB 4;  Length 99;
Best Local Similarity   90.3%;  Pred. No. 7.3e-28;
Matches  56;  Conservative   1;  Mismatches   3;  Indels   2;  Gaps   1;

Qy    12 RRRGCDLSFPSNKLVSGDHCRIVDEKSGGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 71
            |::|--| |||||||||||||||||||||||||||||||||||||||||||||||
Db     9 RDKGPD--FPXNKLVSGDHCRIVDEKSSGVTLEDTSTSGTVINKLKVVKKQTCPLQTGD 66

Qy    72 VI 73
         ||
Db    67 VI 68

RESULT 12
ABU55645
ID   ABU55645 standard; protein; 99 AA.
XX
AC   ABU55645;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polypeptide #732.
XX
KW   Human; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001; 2001US-00764864.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0209963P.
PR   26-JUL-2000; 2000US-0209964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
```

```
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0246617P.
PR   01-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-488783/53.
DR   N-PSDB; AAS26563.
XX
PT   New nucleic acid molecules encoding 461 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives.
XX
PS   Claim 11; SEQ ID NO 1529; 980pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules and their
CC   encoded secreted proteins. The nucleic acids and proteins are used to
CC   prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
```

```
Best Local Similarity   90.3%;  Pred. No. 7.3e-28;
Matches  56;  Conservative   1;  Mismatches   3;  Indels   2;  Gaps   1;

Qy      12  RRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD  71
            | :|||  || |||||||||||||||||||||||||||||||||||||||||||||||||
Db       9  RDKGPD--FPXNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD  66

Qy      72  VI 73
            || |
Db      67  VI 68


RESULT 13
ABU55226
ID   ABU55226 standard; protein; 99 AA.
XX
AC   ABU55226;
XX
DT   18-MAR-2003  (first entry)
XX
DE   Human novel polypeptide #313.
XX
KW   Human; neural disorder; immune system disorder; renal disorder;
KW   muscular disorder; respiratory disease; reproductive disorder;
KW   gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW   hyperproliferative disorder; inflammatory disease; allergic reaction;
KW   blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW   cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW   haemostatic; antiarteriosclerotic.
XX
OS   Homo sapiens.
XX
PN   US2002132753-A1.
XX
PD   19-SEP-2002.
XX
PF   17-JAN-2001;  2001US-00764864.
XX
PR   31-JAN-2000;  2000US-0179065P.
PR   04-FEB-2000;  2000US-0180628P.
PR   28-JUN-2000;  2000US-0214886P.
PR   07-JUL-2000;  2000US-0216647P.
PR   07-JUL-2000;  2000US-0216880P.
PR   11-JUL-2000;  2000US-0217487P.
PR   11-JUL-2000;  2000US-0217496P.
PR   14-JUL-2000;  2000US-0218290P.
PR   26-JUL-2000;  2000US-0220963P.
PR   26-JUL-2000;  2000US-0220964P.
PR   14-AUG-2000;  2000US-0224518P.
PR   14-AUG-2000;  2000US-0224519P.
PR   14-AUG-2000;  2000US-0225267P.
PR   14-AUG-2000;  2000US-0225268P.
PR   14-AUG-2000;  2000US-0225270P.
PR   14-AUG-2000;  2000US-0225447P.
PR   14-AUG-2000;  2000US-0225757P.
PR   14-AUG-2000;  2000US-0225758P.
PR   22-AUG-2000;  2000US-0226868P.
PR   30-AUG-2000;  2000US-0228924P.
PR   01-SEP-2000;  2000US-0229287P.
PR   01-SEP-2000;  2000US-0229343P.
PR   01-SEP-2000;  2000US-0229344P.
PR   01-SEP-2000;  2000US-0229345P.
PR   05-SEP-2000;  2000US-0229509P.
PR   05-SEP-2000;  2000US-0229513P.
PR   08-SEP-2000;  2000US-0231413P.
PR   21-SEP-2000;  2000US-0234223P.
PR   21-SEP-2000;  2000US-0234274P.
PR   25-SEP-2000;  2000US-0234997P.
PR   27-SEP-2000;  2000US-0235834P.
PR   29-SEP-2000;  2000US-0236327P.
PR   29-SEP-2000;  2000US-0236368P.
PR   29-SEP-2000;  2000US-0236369P.
PR   29-SEP-2000;  2000US-0236369P.
PR   29-SEP-2000;  2000US-0236370P.
PR   02-OCT-2000;  2000US-0236802P.
PR   02-OCT-2000;  2000US-0237037P.
PR   02-OCT-2000;  2000US-0237038P.
PR   02-OCT-2000;  2000US-0237039P.
PR   02-OCT-2000;  2000US-0237040P.
PR   13-OCT-2000;  2000US-0239935P.
PR   20-OCT-2000;  2000US-0240960P.
PR   20-OCT-2000;  2000US-0241785P.
PR   20-OCT-2000;  2000US-0241809P.
PR   01-NOV-2000;  2000US-0244617P.
PR   17-NOV-2000;  2000US-0249299P.
PR   08-DEC-2000;  2000US-0251856P.
PR   08-DEC-2000;  2000US-0251868P.
PR   08-DEC-2000;  2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA, Ruben SM, Barash SC;
XX
DR   WPI; 2003-147444/14.
DR   N-PSDB; ABX73904.
XX
PT   New polypeptides and nucleic acids, useful in gene therapy for treating,
PT   inhibiting or preventing e.g. neural, immune system, muscular,
PT   respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT   renal disorders.
XX
PS   Claim 11; SEQ ID NO 1529; 402pp; English.
XX
CC   The invention relates to human novel polypeptides and their associated
CC   polynucleotides. The polypeptides and polynucleotides are useful in gene
CC   therapy for treating, inhibiting or preventing neural disorders, immune
CC   system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC   and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC   nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC   gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC   (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC   heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC   renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC   leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC   appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC   related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC   infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC   ABU55748 represent human novel polypeptides of the invention
XX
SQ   Sequence 99 AA;

Query Match         72.8%;  Score 275;  DB 6;  Length 99;
```

PR 29-SEP-2000; 2000US-0363370P.
PR 02-OCT-2000; 2000US-0366802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
DR WPI; 2003-147444/14.
DR N-PSDB; ABX73485.
XX
PT New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.
XX
PS Claim 11; SEQ ID NO 1110; 402pp; English.
XX
CC The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC ABU55748 represent human novel polypeptides of the invention
XX
SQ Sequence 99 AA;

Query Match　　　　72.8%;　　Score 275;　DB 6;　Length 99;
Best Local Similarity　90.3%;　Pred. No. 7.3e-28;
Matches　56;　Conservative　1;　Mismatches　3;　Indels　2;　Gaps　1;

Qy　　12　RRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD　71
　　　　　|：||　|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　9　RDKGPD--FPXNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQTGD　66

Qy　　72　VI　73
　　　　　||
Db　　67　VI　68

RESULT 14
AAW73894
ID AAW73894 standard; protein; 776 AA.
XX
AC AAW73894;
XX
DT 08-APR-1999 (first entry)
XX
DE Yeast MEC2 protein sequence.
XX

KW Cell cycle checkpoint gene; yeast; radiation resistance; detection;
KW G2/M checkpoint.
XX
OS Saccharomyces cerevisiae.
XX
PN US5866338-A.
XX
PD 02-FEB-1999.
XX
PF 06-JUN-1997; 97US-00870693.
XX
PR 12-MAY-1992; 92US-00882051.
PR 14-MAY-1992; 92US-00884426.
PR 12-MAY-1993; 93WO-US004458.
PR 18-FEB-1994; 94US-0019846.
XX
PA (UNIW) UNIV WASHINGTON.
PA (UYAR-) UNIV ARIZONA.
PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI Plon SE, Groudine MT, Hartwell LH, Weinert TA;
XX
DR WPI; 1999-141919/12.
DR N-PSDB; AAX01275.
XX
PT Nucleotide sequences that hybridise to huRAD cDNA - method for isolating
PT human checkpoint cDNA, and antibody to CDC34.
XX
PS Example 4; Col 89-96; 73pp; English.
XX
CC This sequence is the yeast MEC2 protein. The DNA encoding this sequence
CC is necessary for G2/M cell cycle checkpoint control. This sequence was
CC identified using the method of the invention for isolating a human
CC checkpoint cDNA that is capable of restoring growth at a restrictive
CC temperature in a yeast test cell, where the yeast test cell comprises a
CC genome having a first gene that forms a DNA strand break at a restrictive
CC temperature and a second gene that fails to induce a cell cycle arrest in
CC response to the DNA strand break, whereby the growth of the yeast test
CC cell is inhibited at the restrictive temperature, comprising: (a)
CC obtaining a human cDNA library comprising several human cDNA clones; (b)
CC inserting the human cDNA clones individually into plasmid vectors
CC containing a selectable marker gene; (c) transforming a culture of the
CC yeast test cells with the plasmid vectors from the preceding step; (d)
CC selecting for yeast test cells transformed with the selectable marker
CC gene; (e) growing the selected transformants at the restrictive
CC temperature and isolating a candidate transformant capable of growing at
CC the restrictive temperature; and (f) identifying the human cDNA carried
CC by the candidate transformant as a human checkpoint cDNA by sequencing
CC the human cDNA carried by the candidate transformant and determining that
CC the human cDNA is less than 50% homologous with both the first gene and
CC the second gene. The nucleotide sequence of huRADcompB can be used to
CC confer radiation resistance on a cell
XX
SQ Sequence 776 AA;

Query Match　　　　28.4%;　　Score 107.5;　DB 2;　Length 776;
Best Local Similarity　37.7%;　Pred. No. 8.7e-05;
Matches　26;　Conservative　9;　Mismatches　33;　Indels　1;　Gaps　1;

Qy　　5　KREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQT　64
　　　　　|：||||||||　||||||：||：：：|||||||||||||：|||||||||：||　|||||
Db　　63　KKVWTFGRNPACDYHLGNISRLSNKHPQILLGE-DGNLLLNDISTNGTWLNGQCKVEKNSN　121

Qy　　65　CPLQTGDVI　73
　　　　　|||||
Db　　122　QLLSQGDEI　130

RESULT 15
AAW26664
ID AAW26664 standard; protein; 821 AA.
XX
AC AAW26664;

Search completed: May 7, 2004, 14:45:46
Job time : 17.2046 secs

```
XX
DT   25-MAR-2003   (revised)
DT   25-FEB-1998   (first entry)
XX
DE   Yeast checkpoint control protein MEC2.
XX
KW   MEC2; cell cycle; checkpoint gene; yeast; DNA damage; cancer; therapy.
XX
OS   Saccharomyces cerevisiae.
XX
FH   Key           Location/Qualifiers
FT   Protein       1..776
FT                 /note= "in the Sequence Listing MEC2 terminates at
FT                 residue 775"
XX
PN   US5674996-A.
XX
PD   07-OCT-1997.
XX
PF   18-FEB-1994;   94US-00198446.
XX
PR   12-MAY-1992;   92US-00882051.
PR   14-MAY-1992;   92US-00884426.
PR   12-MAY-1993;   93WO-US004458.
XX
PA   (UYAR-) UNIV ARIZONA.
PA   (UNIW ) UNIV WASHINGTON.
PA   (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI   Groudine MT, Weinert TA, Plon SE, Hartwell LH;
XX
DR   WPI; 1997-502392/46.
DR   N-PSDB; AAT91040.
XX
PT   Nucleotide sequence capable of hybridising with huCDC34 - which is human
PT   checkpoint gene, useful to increase sensitivity of tumour cells to
PT   chemotherapeutic drugs or radiation.
XX
PS   Example 4; Col 85-90; 54pp; English.
XX
CC   This protein is encoded by novel yeast checkpoint control gene MEC2 (see
CC   AAT91040). Yeast checkpoint control genes RAD17, RAD24, MEC1, MEC2 and
CC   MEC3 (AAT91037-41, respectively) are responsible for recognising if the
CC   cell has suffered DNA damage in the form of radiation or chemical damage
CC   or if the cell has failed to complete DNA replication because of chemical
CC   inhibition or intrinsic error. Upon recognising damage or failure, the
CC   genes are responsible for inhibiting mitosis. The purpose of this
CC   checkpoint control is that it preserves the viability of the cell and the
CC   integrity of the genome by providing the cell time to repair these
CC   insults prior to undertaking mitosis. The genes are potentially useful in
CC   developing cancer chemotherapeutics, cancer chemoprevention agents, and
CC   environmental toxicology tests. They can be used to produce proteins (see
CC   AAW26661-65) that can then be screened for chemical agents that would
CC   interfere with checkpoint controls. Cloned genes can also be used to
CC   develop yeast strains in which these genes are deleted. Such yeast
CC   strains can then be used to find the homologous human genes (see AAT91034
CC   -36). (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 821 AA;
```

Query Match   28.4%;   Score 107.5;   DB 2;   Length 821;
Best Local Similarity   37.7%;   Pred. No. 9.3e-05;
Matches 26;   Conservative 9;   Mismatches 33;   Indels   1;   Gaps   1;

```
Qy     5  KREWTIGRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQT 64
           : :| |||  ||       ||        : |||  |||: :||  |  || |:|  |||
Db    63  KKVWTFGRNPACDYHLGNISRLSNKFQILLGE-DGNLILNDISTNGTWLNGQKVEKNSN 121

Qy    65  CPLQTGDVI 73
           |  | |||
Db   122  QLLSQGDEI 130
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 14:42:23 ; Search time 4.85639 Seconds
                                    (without alignments)
                                    776.028 Million cell updates/sec

Title: US-10-048-046-2_COPY_31_103
Perfect score: 378
Sequence: 1 VLLRKREWTIGRRRGCDLSF.........INKLKVVKKQTCPLQTGDVI 73

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued Patents AA:*
          1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
          2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
          3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
          4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
          5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
          6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	107.5	28.4	776	1	US-08-198-446B-17	Sequence 17, Appl
2	107.5	28.4	776	2	US-08-870-693-17	Sequence 17, Appl
3	107.5	28.4	821	1	US-09-198-446B-6	Sequence 6, Appli
4	107.5	28.4	821	4	US-08-870-693-6	Sequence 6, Appli
5	97	25.7	65	4	US-09-300-008B-46	Sequence 46, Appl
6	79.5	21.0	558	4	US-09-252-991A-17202	Sequence 17202, A
7	76.5	20.2	517	4	US-09-849-617-2	Sequence 2, Appli
8	74.5	19.7	64	4	US-09-300-008B-44	Sequence 44, Appl
9	70.5	18.7	877	4	US-09-328-352-8162	Sequence 8162, Ap
10	68	18.0	754	4	US-09-300-008B-2	Sequence 2, Appli
11	67	17.7	63	4	US-09-300-008B-43	Sequence 43, Appl
12	64.5	17.1	64	4	US-09-300-008B-41	Sequence 41, Appl
13	61	16.1	212	4	US-09-328-352-5652	Sequence 5652, Ap
14	61	16.1	1070	4	US-09-107-532A-6735	Sequence 6735, Ap
15	59.5	15.7	804	3	US-08-855-910-4	Sequence 4, Appli
16	59.5	15.7	824	4	US-09-134-000C-4908	Sequence 4908, Ap
17	59	15.6	1045	1	US-07-596-467-6	Sequence 6, Appli
18	59	15.6	1045	1	US-07-934-374-6	Sequence 6, Appli
19	59	15.6	1045	1	US-07-783-861C-6	Sequence 6, Appli
20	58.5	15.5	476	4	US-09-252-991A-22693	Sequence 22693, A
21	58.5	15.5	502	4	US-09-252-991A-30007	Sequence 30007, A
22	58	15.3	721	4	US-08-851-435-2	Sequence 2, Appli
23	58	15.3	746	4	US-08-851-435-6	Sequence 6, Appli
24	57.5	15.2	80	4	US-09-540-236-2410	Sequence 2410, Ap
25	57.5	15.2	443	4	US-09-134-001C-3148	Sequence 3148, Ap
26	57.5	15.2	632	3	US-09-354-129-8	Sequence 8, Appli
27	57.5	15.2	632	4	US-09-504-357-8	Sequence 8, Appli
28	57	15.1	368	3	US-08-961-083-76	Sequence 76, Appl
29	57	15.1	368	4	US-09-536-784-76	Sequence 76, Appl
30	56.5	15.1	870	4	US-09-543-681A-5184	Sequence 5184, Ap
31	56.5	14.9	440	4	US-09-198-452A-755	Sequence 755, App
32	56	14.8	543	4	US-09-529-093A-2	Sequence 2, Appli
33	56	14.8	543	4	US-09-529-154-2	Sequence 2, Appli
34	56	14.8	830	3	US-08-804-439A-20	Sequence 20, Appl
35	56	14.8	830	3	US-08-720-229-20	Sequence 20, Appl
36	56	14.8	878	4	US-09-489-039A-13174	Sequence 13174, A
37	56	14.8	913	4	US-09-252-991A-22127	Sequence 22127, A
38	56	14.8	3256	4	US-09-919-172-98	Sequence 98, Appl
39	56	14.7	3256	4	US-09-976-594-22	Sequence 22, Appl
40	55.5	14.7	1548	4	US-09-252-991A-22301	Sequence 22301, A
41	55	14.6	65	4	US-09-300-008B-48	Sequence 48, Appl
42	55	14.6	283	4	US-09-107-532A-6481	Sequence 6481, Ap
43	55	14.6	312	4	US-09-107-532A-5790	Sequence 5790, Ap
44	55	14.6	356	4	US-09-540-236-3075	Sequence 3075, Ap
45	55	14.6	1302	4	US-09-423-890-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-198-446B-17
; Sequence 17, Application US/08198446B
; Patent No. 5674996
; GENERAL INFORMATION:
; APPLICANT: Hartwell, Leland H.
; APPLICANT: Weinert, Ted A.
; APPLICANT: Plon, Sharon E.
; APPLICANT: Groudine, Mark T.
; TITLE OF INVENTION: Cell Cycle Checkpoint Genes
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/198,446B
; FILING DATE: 18-FEB-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCR17537
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100
; TELEFAX: 206-224-0779
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 776 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Yeast MEC2 protein
US-08-198-446B-17

Query Match   28.4%; Score 107.5; DB 1; Length 776;
Best Local Similarity   37.7%; Pred. No. 2.4e-05;
Matches   26; Conservative   9; Mismatches 33; Indels   1; Gaps   1;

Qy   5 KREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQT 64
            |.|| ||||| : ..|:: .|::: | .:|::| : :|::::.::|:||.:|.||||
Db   63 KKVWTFGRNPACDYHLGNISRLSNKHFQILLGE-DGNLLNDISTNGTWLNGQKVEKNSN 121

```
Qy     65  CPLQTGDVI  73
              ||||||
Db    122  QLLSQGDEI 130

RESULT 2
US-08-870-693-17
; Sequence 17, Application US/08870693
; Patent No. 5866338
; GENERAL INFORMATION:
;   APPLICANT: Hartwell, Leland H.
;   APPLICANT: Weinert, Ted A.
;   APPLICANT: Plon, Sharon E.
;   APPLICANT: Groudine, Mark T.
;   TITLE OF INVENTION: Cell Cycle Checkpoint Genes
;   NUMBER OF SEQUENCES: 19
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
;   STREET: 1420 Fifth Ave., Suite 2800
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98101-2347
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/870,693
;   FILING DATE:
;   CLASSIFICATION: 536
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/198,446
;   FILING DATE: February 18, 1994
;   APPLICATION NUMBER: PCT/US93/04458
;   FILING DATE: May 12, 1993
;   APPLICATION NUMBER: US 07/884,426
;   FILING DATE: May 14, 1992
;   APPLICATION NUMBER: US 07/882,051
;   FILING DATE: May 12, 1992
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Sheiness, Diana K.
;   REGISTRATION NUMBER: 35,356
;   REFERENCE/DOCKET NUMBER: FHCR110798
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-682-8100
;   TELEFAX: 206-224-0779
; INFORMATION FOR SEQ ID NO: 17:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 776 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   DESCRIPTION: yeast MEC2 protein
US-08-870-693-17

Query Match       28.4%;  Score 107.5;  DB 2;  Length 776;
Best Local Similarity  37.7%;  Pred. No. 2.4e-05;
Matches   26;  Conservative   9;  Mismatches  33;  Indels   1;  Gaps   1;

Qy      5  KREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQT 64
              |:||||    :|||  :|::|      |:|:  |    |    |  |: : |||
Db     63  KKVWTFGRNPACDYHLGNISRLSNKHFQILLGE-DGNLLLNDISTNGTWLNGQKVEKNSN 121

Qy     65  CPIQTGDVI  73
              ||||||
Db    122  QLLSQGDEI 130

RESULT 3
US-08-198-446B-6
; Sequence 6, Application US/08198446B
; Patent No. 5674996
; GENERAL INFORMATION:
;   APPLICANT: Hartwell, Leland H.
;   APPLICANT: Weinert, Ted A.
;   APPLICANT: Plon, Sharon E.
;   APPLICANT: Groudine, Mark T.
;   TITLE OF INVENTION: Cell Cycle Checkpoint Genes
;   NUMBER OF SEQUENCES: 19
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
;   STREET: 1420 Fifth Ave., Suite 2800
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98101-2347
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/198,446B
;   FILING DATE: 18-FEB-1994
;   CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Sheiness, Diana K.
;   REGISTRATION NUMBER: 35,356
;   REFERENCE/DOCKET NUMBER: FHCR17537
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 206-682-8100
;   TELEFAX: 206-224-0779
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 821 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   DESCRIPTION: yeast MEC2 protein
US-08-198-446B-6

Query Match       28.4%;  Score 107.5;  DB 1;  Length 821;
Best Local Similarity  37.7%;  Pred. No.2.6e-05;
Matches   26;  Conservative   9;  Mismatches  33;  Indels   1;  Gaps   1;

Qy      5  KREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQT 64
              |:||||    :|||  :|::|      |:|:  |    |    |  |: : |||
Db     63  KKVWTFGRNPACDYHLGNISRLSNKHFQILLGE-DGNLLLNDISTNGTWLNGQKVEKNSN 121

Qy     65  CPLQTGDVI  73
              ||||||
Db    122  QLLSQGDEI 130

RESULT 4
US-08-870-693-6
; Sequence 6, Application US/08870693
; Patent No. 5866338
; GENERAL INFORMATION:
;   APPLICANT: Hartwell, Leland H.
;   APPLICANT: Weinert, Ted A.
;   APPLICANT: Plon, Sharon E.
;   APPLICANT: Groudine, Mark T.
;   TITLE OF INVENTION: Cell Cycle Checkpoint Genes
;   NUMBER OF SEQUENCES: 19
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
;   STREET: 1420 Fifth Ave., Suite 2800
;   CITY: Seattle
;   STATE: WA
;   COUNTRY: USA
;   ZIP: 98101-2347
```

```
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:    Floppy disk
;     COMPUTER:       IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:       PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/870,693
;     FILING DATE:
;     CLASSIFICATION:  536
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/198,446
;     FILING DATE:    February 18, 1994
;     APPLICATION NUMBER:  PCT/US93/04458
;     FILING DATE:    May 12, 1993
;     APPLICATION NUMBER:  US 07/884,426
;     FILING DATE:    May 14, 1992
;     APPLICATION NUMBER:  US 07/882,051
;     FILING DATE:    May 12, 1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Shelness, Diana K.
;     REGISTRATION NUMBER:  35,356
;     REFERENCE/DOCKET NUMBER:  FHCR110798
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  206-682-8100
;     TELEFAX:  206-224-0779
;   INFORMATION FOR SEQ ID NO:  6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  821 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
;   DESCRIPTION:  yeast MEC2 protein
US-08-870-693-6

Query Match             28.4%;   Score 107.5;  DB 2;   Length 821;
Best Local Similarity   37.7%;   Pred. No. 2.6e-05;   Indels  33;   Gaps  1;
Matches   26;   Conservative   9;   Mismatches   33;   Indels   33;   Gaps  1;

Qy    5  KREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQT  64
            | |||||   |   |:::   :|||||||||||||||| ||||||:|   |  ||
Db   63  KKVWTFGRNPACDYHLGNISRLSNKHFQILLGE-DGNLLLNDISTNGTWLNGQKVEKNSN  121

Qy   65  CPLQTGDVI  73
            |||||||
Db  122  QLLSQGDEI  130

RESULT 5
US-09-300-008B-46
;  Sequence 46, Application US/09300008B
;  Patent No. 6458534
;  GENERAL INFORMATION:
;    APPLICANT:  Concannon et al.
;    TITLE OF INVENTION:  A GENE ASSOCIATED WITH NIJMEGEN BREAKAGE
;    TITLE OF INVENTION:  SYNDROME, ITS GENE PRODUCT AND METHODS FOR THEIR USE
;    FILE REFERENCE:  9924-0003-228
;    CURRENT APPLICATION NUMBER:  US/09/300,008B
;    CURRENT FILING DATE:  1999-04-27
;    PRIOR APPLICATION NUMBER:  US 60/083,269
;    PRIOR FILING DATE:  1998-04-27
;    SOFTWARE:  FastSEQ for Windows Version 3.0
;    NUMBER OF SEQ ID NOS:  64
;    SEQ ID NO 46
;      LENGTH:  65
;      TYPE:  PRT
;      ORGANISM:  Saccharomyces cerevisiae
US-09-300-008B-46

Query Match             25.7%;   Score 97;   DB 4;   Length 65;
Best Local Similarity   36.4%;   Pred. No. 2.3e-05;   Indels  30;   Gaps  1;
Matches   24;   Conservative   8;   Mismatches   30;   Indels   30;   Gaps  1;
```

```
Qy    8  WTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPL  67
            ||||||   :|   |:::   :|||||||||||||||| |||||::|  | :||| |
Db    1  WTFGRNPACDYHLGNISRLSNKHFQILL----GNLLLNDISTNGTWLNGQKVEKNSNQLL  56

Qy   68  QTGDVI  73
            ||||
Db   57  SQGDEI  62

RESULT 6
US-09-252-991A-17202
;  Sequence 17202, Application US/09252991A
;  Patent No. 6551795
;  GENERAL INFORMATION:
;    APPLICANT:  Rubenfield et al.
;    TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;    TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;    FILE REFERENCE:  107196.136
;    CURRENT APPLICATION NUMBER:  US/09/252,991A
;    CURRENT FILING DATE:  1999-02-18
;    PRIOR APPLICATION NUMBER:  US 60/074,788
;    PRIOR FILING DATE:  1998-02-18
;    PRIOR APPLICATION NUMBER:  US 60/094,190
;    PRIOR FILING DATE:  1998-07-27
;    NUMBER OF SEQ ID NOS:  33142
;    SEQ ID NO 17202
;      LENGTH:  558
;      TYPE:  PRT
;      ORGANISM:  Pseudomonas aeruginosa
US-09-252-991A-17202

Query Match             21.0%;   Score 79.5;   DB 4;   Length 558;
Best Local Similarity   32.5%;   Pred. No. 0.07;   Indels  27;   Gaps  4;
Matches   25;   Conservative   16;   Mismatches   27;   Indels   27;   Gaps  4;

Qy    2  LLRKREWTIGRRRGCDLSFPS-NKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVI----NK  56
            :|:|  ||||||   |   :||||||||   :  :   ::   :
Db   85  VLDQCQLITGRGPDNDWVLPDPERLVSSRHCTIL--NRDGVYYLTDTSTNGVLLVNAGHR  142

Qy   57  LKVVKKQTCPLQTGDVI  73
            |       ||  ||
Db  143  LR--RGNSEPLQDGETV  157

RESULT 7
US-09-849-617-2
;  Sequence 2, Application US/09849617
;  Patent No. 6593110
;  GENERAL INFORMATION:
;    APPLICANT:  CALIFORNIA INSTITUTE OF TECHNOLOGY
;    APPLICANT:  GUO, Zilian
;    APPLICANT:  DUNPHY, William
;    TITLE OF INVENTION:  CHECKPOINT-ACTIVATING OLIGONUCLEOTIDES
;    FILE REFERENCE:  CIT1350-1
;    CURRENT APPLICATION NUMBER:  US/09/849,617
;    CURRENT FILING DATE:  2001-09-17
;    PRIOR APPLICATION NUMBER:  US 60/202,028
;    PRIOR FILING DATE:  2000-05-04
;    NUMBER OF SEQ ID NOS:  14
;    SOFTWARE:  PatentIn version 3.0
;    SEQ ID NO 2
;      LENGTH:  517
;      TYPE:  PRT
;      ORGANISM:  Xenopus laevis
US-09-849-617-2

Query Match             20.2%;   Score 76.5;   DB 4;   Length 517;
Best Local Similarity   31.9%;   Pred. No. 0.16;   Indels  29;   Gaps  2;
Matches   23;   Conservative   9;   Mismatches   29;   Indels   29;   Gaps  2;

Qy    7  EWTIGRRRGCDLSF-------PSNKLVSGDHCRIVVDEKSGQ---VTLEDTSTSGTVIN  55
            | :|:|||||   |       | :||||||||   ||   |      :|||||||| |
Db   84  EYVFGRDKKCCDYTFDIPVLNQTDRYKTYSKRHFRIFQELGHGHSRVANIEDLSGNGTFVN  143
```

```
Qy    56  KLKVVKKQTCPL  67
          | :||  :||||
Db   144  KEIIGKGRTLPL  155

RESULT 8
US-09-300-008B-44
; Sequence 44, Application US/09300008B
; Patent No. 6458534
; GENERAL INFORMATION:
; APPLICANT: Concannon et al.
; TITLE OF INVENTION: A GENE ASSOCIATED WITH NIJMEGEN BREAKAGE
; TITLE OF INVENTION: SYNDROME, ITS GENE PRODUCT AND METHODS FOR THEIR USE
; FILE REFERENCE: 9924-0003-228
; CURRENT APPLICATION NUMBER: US/09/300,008B
; CURRENT FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: US 60/083,269
; PRIOR FILING DATE: 1998-04-27
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-09-300-008B-44

Query Match           19.7%;   Score 74.5;  DB 4;  Length 64;
Best Local Similarity 31.8%;   Pred. No. 0.019;
Matches    21;  Conservative   9;  Mismatches  31;  Indels   5;  Gaps   2;

Qy     8  WTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPL  67
          |  |::                 |||  :|  :|  ||||||||  ::
Db     1  WGFGRHKSCEVVL-NGPRVSNFHFEIY----QGVVFLHDHSSNGTFLNFERLAKNSRTIL  55

Qy    68  QTGDVI  73
          ||||
Db    56  SNGDEI  61

RESULT 9
US-09-328-352-8162
; Sequence 8162, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8162
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-8162

Query Match           18.7%;   Score 70.5;  DB 4;  Length 877;
Best Local Similarity 25.9%;   Pred. No. 1.9;
Matches    22;  Conservative  12;  Mismatches  24;  Indels  27;  Gaps   4;

Qy     1  VLLRKREWTIGRRR--GCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTST--------  49
          |    ||:|:|:| |:              :|:||   ||||::
Db   430  VQFRLRDWGVSRQRYWGCPI-----PMINCDTC-------GQVTVPEDQLPVVLPTDVV  476

Qy    50  ---SGTVINKLKVVKKQTCPLQTGD  71
              ||:  ||| |:|||  ||| |
Db   477  PDGSGNPLNKMPEFYETKCPCCGGD  501

RESULT 10
US-09-300-008B-2
```

```
; Sequence 2, Application US/09300008B
; Patent No. 6458534
; GENERAL INFORMATION:
; APPLICANT: Concannon et al.
; TITLE OF INVENTION: A GENE ASSOCIATED WITH NIJMEGEN BREAKAGE
; TITLE OF INVENTION: SYNDROME, ITS GENE PRODUCT AND METHODS FOR THEIR USE
; FILE REFERENCE: 9924-0003-228
; CURRENT APPLICATION NUMBER: US/09/300,008B
; CURRENT FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: US 60/083,269
; PRIOR FILING DATE: 1998-04-27
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 754
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-300-008B-2

Query Match           18.0%;   Score 68;  DB 4;  Length 754;
Best Local Similarity 24.7%;   Pred. No. 3.2;
Matches    20;  Conservative  21;  Mismatches  30;  Indels  10;  Gaps   2;

Qy     2  LLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVD-------EKSGQVTLEDTSTSGT  52
          |  |  ||  :|| ::|||  |  : :|:|  : :|       :||||:|||||||||
Db    18  LLTGVEYVVG-RKNCAILIENDQSISRNHAVLTANFSVTNLSQTDEIPVLTLKDNSKYGT  76

Qy    53  VINKLKVVKKQTCPLQTGDVI  73
          :|| :|       :||||||||
Db    77  FVNEEKMQNGFSRTLKSGDGI  97

RESULT 11
US-09-300-008B-43
; Sequence 43, Application US/09300008B
; Patent No. 6458534
; GENERAL INFORMATION:
; APPLICANT: Concannon et al.
; TITLE OF INVENTION: A GENE ASSOCIATED WITH NIJMEGEN BREAKAGE
; TITLE OF INVENTION: SYNDROME, ITS GENE PRODUCT AND METHODS FOR THEIR USE
; FILE REFERENCE: 9924-0003-228
; CURRENT APPLICATION NUMBER: US/09/300,008B
; CURRENT FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: US 60/083,269
; PRIOR FILING DATE: 1998-04-27
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 43
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-300-008B-43

Query Match           17.7%;   Score 67;  DB 4;  Length 63;
Best Local Similarity 34.3%;   Pred. No. 0.18;
Matches    23;  Conservative  10;  Mismatches  24;  Indels  10;  Gaps   3;

Qy     9  TIGRRRGCD--LSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQTCP  66
          ||||:|||     ::|| :||   ::|:|:|   :||||| :|| ||| |
Db     2  TIGRSRSCDVILSEPDISTFHAEFHLLLIN------VIDKSRNGTFINGNRLVKKDYI-  53

Qy    67  LQTGDVI  73
          |:||||
Db    54  LKNGDRI  60

RESULT 12
US-09-300-008B-41
; Sequence 41, Application US/09300008B
; Patent No. 6458534
; GENERAL INFORMATION:
; APPLICANT: Concannon et al.
; TITLE OF INVENTION: A GENE ASSOCIATED WITH NIJMEGEN BREAKAGE
```

~ TITLE OF INVENTION: SYNDROME, ITS GENE PRODUCT AND METHODS FOR THEIR USE
~ FILE REFERENCE: 9924-0003-228
~ CURRENT APPLICATION NUMBER: US/09/300,008B
~ CURRENT FILING DATE: 1999-04-27
~ PRIOR APPLICATION NUMBER: US 60/083,269
~ PRIOR FILING DATE: 1998-04-27
~ NUMBER OF SEQ ID NOS: 64
~ SOFTWARE: FastSEQ for Windows Version 3.0
~ SEQ ID NO 41
~ LENGTH: 64
~ TYPE: PRT
~ ORGANISM: Homo sapien
US-09-300-008B-41

Query Match 17.1%; Score 64.5; DB 4; Length 64;
Best Local Similarity 25.8%; Pred. No. 0.38;
Matches 17; Conservative 19; Mismatches 25; Indels 5; Gaps 2;

```
Qy   8 WTIGRRRGCDLSFPSNKLVSGDHCRIVDEKSGGVTLEDTSTSGTVINKLKVVKKQTCPL 67
         :  :|  |:   :: :::|: |  :       | |: || |:| :||::|:::|
Db   1 YVVG-RKNCAILIENDQSISRNHAVLT---ANVLTLKDNSKYGTFVNEEKMQNGFSRTL 55

Qy  68 QTGDVI 73
         :|||
Db  56 KSGDGI 61
```

RESULT 13
US-09-328-352-5652
; Sequence 5652, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5652
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5652

Query Match 16.1%; Score 61; DB 4; Length 212;
Best Local Similarity 23.2%; Pred. No. 5.1;
Matches 16; Conservative 24; Mismatches 23; Indels 6; Gaps 4;

```
Qy   6 REWTIGRRRGCDLSFPSNKLVSGDHCRIVDEKSGGVTLED-TSTSGTVINKLKVVKKQT 64
        |  :||: | :|   :  :::  ::|:  :|   |:||| :| :: |  |
Db  24 RDMLVGRHQDADLILQAAE-ISRRHAALLLKDQA--LMVQDLNSSNGTFVNDMRI--EQE 78

Qy  65 CPLQTGDVI 73
         |::  ||:
Db  79 KQLHDGDIV 87
```

RESULT 14
US-09-107-532A-6735
; Sequence 6735, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA

; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781) 893-5007
; TELEFAX: (781) 893-8277
; INFORMATION FOR SEQ ID NO: 6735:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...1070
; SEQUENCE DESCRIPTION: SEQ ID NO: 6735:
US-09-107-532A-6735

Query Match 16.1%; Score 61; DB 4; Length 1070;
Best Local Similarity 42.3%; Pred. No. 41;
Matches 11; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

```
Qy   18 LSFPSNKLVSGDHCRIVDEKSGQVT 43
          ||:|: :| :|:  ::|:|:::::|
Db  866 LSIPENQLFAGDVLEIVLNDQAGEAT 891
```

RESULT 15
US-08-855-910-4
; Sequence 4, Application US/08855910
; Patent No. 6221640
; GENERAL INFORMATION:
; APPLICANT: Tao, Jianshi
; APPLICANT: Sassanfar, Mandana
; APPLICANT: Gallant, Paul L.
; APPLICANT: Shen, Xiaoyu
; APPLICANT: Avruch, Anthony S.
; APPLICANT: Yu, Russell V.
; APPLICANT: Nair, Shamila
; TITLE OF INVENTION: ENTEROCOCCAL AMINOACYL-tRNA SYNTHETASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACIDS AND STRAINS COMPRISING SAME
; NUMBER OF SEQUENCES: 71
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```
;       APPLICATION NUMBER:  US/08/855,910
;          FILING DATE:  14-MAY-1997
;          CLASSIFICATION:  435
;       ATTORNEY/AGENT INFORMATION:
;          NAME: Brock, David E.
;          REGISTRATION NUMBER: 22,592
;          REFERENCE/DOCKET NUMBER:  CPI95-08
;       TELECOMMUNICATION INFORMATION:
;          TELEPHONE:   (781) 861-6240
;          TELEFAX:    (781) 861-9540
;    INFORMATION FOR SEQ ID NO:  4:
;       SEQUENCE CHARACTERISTICS:
;          LENGTH:  804 amino acids
;          TYPE:  amino acid
;          TOPOLOGY:  linear
;       MOLECULE TYPE:  protein
US-08-855-910-4

Query Match       15.7%;   Score 59.5;  DB 3;  Length 804;
Best Local Similarity   31.5%;   Pred. No. 45;
Matches   23;  Conservative  12;  Mismatches   21;  Indels   17;  Gaps    4;

Qy      5 KREWTIGRRRGCDLSFPSNKLVSGDHCRIVDEKSGQV--TLEDT---STSGTVINKLKV 59
            :|| |||| | |::|            :: | |__|  || | :| ||  :||::
Db    218 QRNW-IGRSEGANVTF----------KVAGTEESFTVFTTRPDTLFGATYTVLAPELEL 265

Qy     60 VKKQTCPLQTGDV 72
            ||| ||_|||_|
Db    266 VKKITTPEQTAAV 278

Search completed: May  7, 2004, 14:51:22
Job time : 5.85639 secs
```

OM protein - protein search, using sw model

```
          GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:   May 7, 2004, 14:41:17 ; Search time 4.08553 Seconds
                                (without alignments)
                                1718.743 Million cell updates/sec

Title:    US-10-048-046-2_COPY_31_103
Perfect score: 378
Sequence: 1 VLLRKREWTIGRRRGCDLSF.........INKLKVVKKQTCPLQTGDVI 73

Scoring table: BLOSUM62
          Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:    283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : PIR_78:*
           1: Pir1:*
           2: pir2:*
           3: pir3:*
           4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	107.5	28.4	821	1	A39616	protein kinase RAD
2	88	23.3	159	2	A96962	FHA-domain contain
3	79.5	21.0	497	2	F83634	hypothetical prote
4	79	20.9	358	2	A61188	probable transcrip
5	77.5	20.5	460	2	S58882	protein kinase Cds
6	76	20.1	2541	2	T29340	hypothetical prote
7	74.5	19.7	265	2	B61188	SCI protein - mous
8	72	19.0	513	1	S43941	protein kinase DUN
9	71	18.8	409	2	S73738	phosphoglycerate k
10	70.5	18.7	230	2	A83872	hypothetical prote
11	70.5	18.7	301	2	T40402	forkhead nuclear s
12	68.5	18.1	157	2	S76488	conserved hypothet
13	68	18.0	399	2	AC3089	hypothetical prote
14	68	18.0	754	2	G98197	Nijmegen breakage
15	68	18.0	234	2	T00393	hypothetical prote
16	67.5	17.9	546	2	AF2214	adenylate cyclase
17	66.5	17.6	287	2	AE1899	cell wall-binding
18	66	17.5	604	2	B69901	probable membrane
19	66	17.5	692	2	SS2706	PSD-95 binding pro
20	66	17.5	977	2	T00025	DAP-1-alpha protei
21	64	16.9	1441	1	T00014	M polyprotein prec
22	63.5	16.8	445	2	GNVUSV	probable protein k
23	63.5	16.8	938	2	T43420	hypothetical prote
24	63.5	16.8	357	2	C84480	conserved hypothet
25	62.5	16.5	468	2	H82645	probable membrane
26	62.5	16.5	527	2	D96904	hypothetical prote
27	62.5	16.5	273	2	B70700	L-rhamnose operon
28	62	16.4	749	2	AF0041	penicillin-binding
29	62	16.4		2	S75331	
30	62	16.4	1471	2	D72758	hypothetical prote
31	62	16.4	1637	2	T00070	hypothetical prote
32	62	16.4	2938	2	T30249	cell proliferation
33	61.5	16.3	287	2	H64690	type IIS restricti
34	61.5	16.3	463	2	T10015	hypothetical prote
35	61.5	16.3	488	2	F86911	conserved hypothet
36	61	16.1	484	2	S48403	FKH1 protein - yea
37	61	16.1	1433	1	GNVUBW	M polyprotein prec
38	61	16.1	1441	1	GNVULC	M polyprotein prec
39	60.5	16.0	337	2	S75018	adenylate cyclase
40	60.5	16.0	397	2	D83437	hypothetical prote
41	60.5	16.0	661	2	T15073	hypothetical prote
42	60	15.9	353	2	E81680	conserved hypothet
43	60	15.9	560	2	D72478	hypothetical prote
44	60	15.9	952	2	T18837	hypothetical prote
45	60	15.9	969	2	B87083	excinuclease ABC s

ALIGNMENTS

```
RESULT 1
A39616
protein kinase RAD53 (EC 2.7.1.-) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein P2588; protein YPL153c; SPK1 protein
C;Species: Saccharomyces cerevisiae
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 21-Jul-2000
C;Accession: A39616; S65164; S69446; S13321
R;Stern, D.F.; Zheng, P.; Beidler, D.R.; Zerillo, C.
Mol. Cell. Biol. 11, 987-1001, 1991
A;Title: Spk1, a new kinase from Saccharomyces cerevisiae, phosphorylates proteins on se
A;Reference number: A39616; MUID:91117267; PMID:1899289
A;Accession: A39616
A;Molecule type: DNA
A;Residues: 1-821 <STE>
A;Cross-references: GB:M55623; NID:g172656; PIDN:AAA35070.1; PID:g172657
A;Experimental source: strain S288C
R;Zheng, P.; Fay, D.S.; Burton, J.; Xiao, H.; Pinkham, J.L.; Stern, D.F.
Mol. Cell. Biol. 13, 5829-5842, 1993
A;Title: SPK1 is an essential S-phage-specific gene of Saccharomyces cerevisiae that enc
A;Reference number: A54697; MUID:93361015; PMID:8355715
A;Contents: annotation
R;Purnelle, B.; Coster, F.; Goffeau, A.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S65154
A;Accession: S65164
A;Molecule type: DNA
A;Residues: 1-821 <PUR>
A;Cross-references: EMBL:Z73509; NID:g1370325; PIDN:CAA97858.1; PID:g1370326; GSPDB:GN00(
A;Experimental source: strain S288C
R;Purnelle, B.; Comblez, S.; Coster, F.; Naveau, F.; Goffeau, A.
submitted to the EMBL Data Library, March 1996
A;Description: The sequence of 55 kb on the left arm of yeast chromosome XVI identifies ;
ogue to the human phosphotyrosyl phosphatase activator PTPA and a homologue to the plant
A;Reference number: S69428
A;Accession: S69446
A;Molecule type: DNA
A;Residues: 1-821 <PUW>
A;Cross-references: EMBL:X96770; NID:g1403537; PIDN:CAA65568.1; PID:g1403556
C;Genetics:
A;Gene: SGD:RAD53; SPK1; MEC2; SAD1; MIPS:YPL153c
A;Cross-references: SGD:S0006074; MIPS:YPL153c
A;Map position: 16L
C;Function:
A;Description: serine/threonine-specific protein kinase
A;Note: contains low activity as tyrosine-specific protein kinase
C;Superfamily: protein kinase SPK1; kinase interaction domain homology; protein kinase h
C;Keywords: ATP; cell cycle control; nucleus; phosphotransferase; serine/threonine-speci
F;68-133/Domain: kinase interaction domain homology <KIH>
F;196-466/Domain: serine/threonine protein kinase homology <KIN>
F;204-213/Region: protein kinase ATP-binding motif
F;227/Active site: Lys #status predicted
```

Query Match   28.4%;   Score 107.5; DB 1;   Length 821;
Best Local Similarity 37.7%;   Pred. No. 0.00028;
Matches   26; Conservative 9; Mismatches 33;   Indels 1; Gaps 1;

Qy   5 KREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVINKLKVVKKQT 64
         |::||| | |:|| ||| :||||:
Db   63 KKVWTFGRNPACDYHLGNISRLSNKHFQILLGE-DGNLLLNDISTNGTWLNGQKVEKNSN 121

Qy   65 CPLQTGDVI 73
         | |||
Db 122 QLLSQGDEI 130

RESULT 2
A96962
FHA-domain containing secreted protein [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: A96962
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: A96962
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-159 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK78484.1; PID:g15023366; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0504

Query Match   23.3%;   Score 88; DB 2;   Length 159;
Best Local Similarity 41.4%;   Pred. No. 0.0086;
Matches   29; Conservative 12; Mismatches 21;   Indels 8; Gaps 5;

Qy   6 REWTIGRRRGCDLSFPSNK-LVSGDHCRIVVDEKSGQVTLED-TSTSGTVINKLKVVKKQ 63
         |:||| | |:| || |: ||| ||| || :|: :| | ||
Db   85 REITIGRKD--DNSIMLNEGYVSGHHARYL--RNNQYILEDLNSTNGTVLNGQKI--KS 138

Qy   64 TCPLQTGDVI 73
         :| |||
Db 139 KAYIKSGDEI 148

RESULT 3
F83634
hypothetical protein PA0081 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: F83634
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: F83634
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-497 <STO>
A;Cross-references: GB:AE004447; GB:AE004091; NID:g9945902; PIDN:AAG03471.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA0081

Query Match   21.0%;   Score 79.5; DB 2;   Length 497;
Best Local Similarity 32.5%;   Pred. No. 0.29;
Matches   25; Conservative 16; Mismatches 27;   Indels 9; Gaps 4;

Qy   2 LLRKREWTIGRRRGCDLSFPS-NKLVSGDHCRIVVDEKSGQVTLEDTSTSGTVI---NK 56
         : ||| ||| ||| ::   :: :: ::   :: ::

Db   22 VLDQGQLTIGRGPDNDWVLPDPERLVSSRHCTIL--NRDGVYLTDTSTNGVLLVNAGHR 79

Qy   57 LKVVKKQTCPLQTGDVI 73

Db   80 LR--RGNSEPLQDGETV 94

RESULT 4
A61188
probable transcription factor SC1 - human
C;Species: Homo sapiens (man)
C;Date: 12-May-1994 #sequence_revision 09-Sep-1994 #text_change 07-May-1999
C;Accession: A61188
R;Ku, D.H.; Chang, C.; Koniecki, J.; Cannizzaro, L.A.; Boghosian-Sell, L.; Alder, H.; Ba
Cell Growth Differ. 2, 179-186, 1991
A;Title: A new growth-regulated complementary DNA with the sequence of a putative trans-
A;Reference number: A61188; MUID:91329275; PMID:1868030
A;Accession: A61188
A;Molecule type: mRNA
A;Residues: 1-358 <KUA>
A;Note: authors translated the codon TAC for residue 19 as Thr
C;Genetics:
A;Gene: SC1
A;Map position: 6p21-22

Query Match   20.9%;   Score 79; DB 2;   Length 358;
Best Local Similarity 27.8%;   Pred. No. 0.23;
Matches   20; Conservative 20; Mismatches 26;   Indels 6; Gaps 4;

Qy   8 WTIGRRRG-CDLSF-PSNK--LVSGDHCRIVVDEKSG--QVTLEDTSGTVINKLKVVK 61

Db   31 YRLGHRADLCDVALRPQQPEPGLISGIHAELHAEPRGDDWRVSLEDHSSQGTLVNNVRLPR 90

Qy   62 KQTCPLQTGDVI 73
         ::: ||
Db   91 GHRLELSDGDLL 102

RESULT 5
S58882
protein kinase Cds1 (EC 2.7.1.-) [validated] - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 31-Dec-2000
C;Accession: S58882; S71846; T41204; T52473
R;Murakami, H.; Okayama, H.
Nature 374, 817-819, 1995
A;Title: A kinase from fission yeast responsible for blocking mitosis in S phase.
A;Reference number: S58882; MUID:95240713; PMID:7723827
A;Accession: S58882
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-460 <MUR1>
A;Cross-references: EMBL:X85040; NID:g794146
R;Murakami, H.
submitted to the EMBL Data Library, March 1995
A;Accession: S71846
A;Molecule type: mRNA
A;Residues: 1-60, 'G', 62-202, 'I', 204-460 <MUR2>
A;Cross-references: EMBL:X85040; NID:g794146; PIDN:CAA59410.1; PID:g794147
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Jimenez Martinez, J.
submitted to the EMBL Data Library, July 1999
A;Accession: T41204
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-202, 'I', 204-237, 'F', 239-460 <WOO>
A;Cross-references: EMBL:AL109736; NID:g5701956; PIDN:g5701956; PID:g5701966; GSPDB:GN
A;Experimental source: strain 972h(-); cosmid c18B5
R;Lindsay, H.D.; Griffiths, D.J.F.; Edwards, R.J.; Christensen, P.U.; Murray, J.M.; Osma
Genes Dev. 12, 382-395, 1998
A;Title: S-phase specific activation of Cds1 kinase defines a subpathway of the checkpoi
A;Reference number: Z26084; MUID:98119835; PMID:9450932

A;Accession: T52473
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-202, 'I', 204-237, 'F', 239-460 <LIN>
A;Cross-references: EMBL:AJ222869; NID:g2689196; PIDN:CAA11019.1; PID:g2689197
C;Genetics:
A;Gene: SPBC18B5.11c; cds1
A;Map position: 3
A;Introns: 86/3; 103/1; 113/3; 142/1; 228/3; 298/3; 368/2
C;Function:
A;Description: EC 2.7.1.-; protein kinase Cds1 [validated, MUID:98119835]; is required t
A;Note: Cds1 is phosphorylated and activated by S-phase arrest and activated by DNA dama
C;Superfamily: protein kinase Cds1; kinase interaction domain homology; protein kinase h
C;Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase
F;62-133/Domain: kinase interaction domain homology <KIN>
F;165-433/Domain: protein kinase homology <KIN>
F;173-181/Region: protein kinase ATP-binding motif

Query Match   20.5%;   Score 77.5;   DB 2; Length 460;
Best Local Similarity 30.6%;   Pred. No. 0.45;
Matches 22;   Conservative 9;   Mismatches 34;   Indels 7;   Gaps 2;

Qy   8 WTIGRRRGCDLSFPSNKLVSGDHCRIV------VDEKSGQVTLEDTSTSGTVINKLKVVK 61
             |||:||| ||::| |:||| | | : |||::|| | ||:||:|||
Db   60 WRFGRHKSCEVVL-NGPRVSNFHFEIYQGHRNDSDESENVVFLHDHSSNGTFLNFERLAK 118

Qy   62 KQTCPLQTGDVI 73
             || | |||
Db   119 NSRTILSNGDEI 130

RESULT 6
T29340
hypothetical protein F21C10.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T29340
R;Du, Z.; Gattung, S.
submitted to the EMBL Data Library, April 1996
A;Description: The sequence of C. elegans cosmid F21C10.
A;Reference number: Z20610
A;Accession: T29340
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2541 <DUZ>
A;Cross-references: EMBL:U55364; PIDN:AAA97973.1; GSPDB:GN00023; CESP:F21C10.7
A;Experimental source: strain Bristol N2; clone F21C10
C;Genetics:
A;Gene: CESP:F21C10.7
A;Map position: 5
A;Introns: 50/3; 91/3; 153/3; 184/3; 215/1; 254/3; 350/1; 392/3; 538/1; 589/3; 625/3; 75

Query Match   20.1%;   Score 76;   DB 2; Length 2541;
Best Local Similarity 31.0%;   Pred. No. 4.3;
Matches 22;   Conservative 13;   Mismatches 24;   Indels 12;   Gaps 3;

Qy   5 KREWTIGRRRGCDLSFPSNKLVSGDHCRIVVDE----KSGQVTLEDTSTSGTVIN--KLK 58
             | ||||| || :| ||: |:||| | || | |||::|| | ||:|| |:|
Db 1265 KEETAVKESERVHLTF-------SGDHCQMIIDKTVPLDTGIYTVRAKNVHGEVANFCQLR 1318

Qy   59 VVKKQTCPLQT 69
             || ||| |
Db 1319 VVPKKQPPPQT 1329

RESULT 7
B61188
SC1 protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 19-Mar-1997 #sequence_revision 13-Mar-1998 #text_change 07-May-1999
C;Accession: B61188
R;Ku, D.H.; Chang, C.; Koniecki, J.; Cannizzaro, L.A.; Boghosian-Sell, L.; Alder, H.; Ba
Cell Growth Differ. 2, 179-186, 1991

A;Title: A new growth-regulated complementary DNA with the sequence of a putative trans-i
A;Reference number: A61188; MUID:91329275; PMID:1868030
A;Accession: B61188
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-265 <KUA>

Query Match   19.7%;   Score 74.5;   DB 2; Length 265;
Best Local Similarity 28.6%;   Pred. No. 0.55;
Matches 18;   Conservative 18;   Mismatches 22;   Indels 5;   Gaps 3;

Qy   16 CDLSF-PSNK--LVSGDHCRIVVDEKSG--QVTLEDTSTSGTVINKLKVVKKQTCPLQTG 70
             |:: | |:| ||:||: :| :|:::| ::: :
Db   40 CDVALRPQQEPGLISGVHAELHAELQGDDWRVSLEDHSSQGTLVNNVRLPRGHRLELSDG 99

Qy   71 DVI 73
             |:|
Db   100 DLL 102

RESULT 8
S43941
protein kinase DUN1 (EC 2.7.1.-) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein D2370; protein YDL101c
C;Species: Saccharomyces cerevisiae
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jun-2000
C;Accession: S43941; S67643; S67418; S72106
R;Zhou, Z.; Elledge, S.J.
Cell 75, 1119-1127, 1993
A;Title: DUN1 encodes a protein kinase that controls the DNA damage response in yeast.
A;Reference number: S43941; MUID:94084787; PMID:8261511
A;Accession: S43941
A;Molecule type: DNA
A;Residues: 1-513 <ZHO>
A;Cross-references: EMBL:L25548; NID:g435616; PIDN:AAA16324.1; PID:g435617
R;Ballesta, J.P.G.; Remacha, M.; Soler-Mira, A.; Jimenez, A.; Garcia-Cantalejo, J.M.; Bot
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67629
A;Accession: S67643
A;Molecule type: DNA
A;Residues: 1-513 <BAL>
A;Cross-references: EMBL:Z74149; NID:g1431139; PIDN:CAA98668.1; PID:g1431140; GSPDB:GN000
A;Experimental source: strain S288C
R;Boskovic, J.; Saiz, J.E.; Soler-Mira, A.; Garcia-Cantalejo, J.; Revuelta, J.L.; Jiminez
submitted to the EMBL Data Library, February 1996
A;Reference number: S67406
A;Accession: S67418
A;Molecule type: DNA
A;Residues: 1-513 <BOS>
A;Cross-references: EMBL:X95644; NID:g1199535; PIDN:CAA64912.1; PID:g1199548
R;Saiz, J.E.; Buitrago, M.J.; Garcia, R.; Revuelta, J.L.; del Rey, F.
Yeast 12, 1077-1084, 1996
A;Title: The sequence of a 20.3 kb DNA fragment from the left arm of Saccharomyces cerevi
A;Reference number: S72094; MUID:97051597; PMID:896274
A;Accession: S72106
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-513 <SAI>
A;Cross-references: EMBL:X95644; NID:g1199535; PIDN:CAA64912.1; PID:g1199548
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996
C;Genetics:
A;Gene: SGD:DUN1; MIPS:YDL101c
A;Cross-references: MIPS:YDL101c; SGD:S0002259
A;Map position: 4L
C;Function:
A;Description: catalyzes the formation of peptidyl-serine-phosphate or peptidyl-threonine
C;Superfamily: protein kinase DUN1; kinase interaction domain homology; protein kinase ho
C;Keywords: ATP; nucleus; phosphoprotein; phosphotransferase; serine/threonine-specific p
F;58-128/Domain: kinase interaction domain homology <KIH>
F;198-480/Domain: protein kinase homology <KIN>
F;206-214/Region: protein kinase ATP-binding motif

Query Match   19.0%;   Score 72;   DB 1; Length 513;

```
Best Local Similarity  35.7%;  Pred. No. 2.2;
Matches   25;  Conservative  10;  Mismatches  29;  Indels   6;  Gaps   3;

Qy     9 TIGRRRGCD--LSFPSNKLVSGDHCRIVVDEKSGQ---VTLEDTSTSGTVINKLKVVKKQ 63
           :||  ||| :||||||||||||||||   :: ||:::|||||||||:|||
Db    57 TIGRSRSCDVILSEPDISTFHAEFHLLQMDVDNFQRNLINVIDKSRNGTFINGNRLLVKKD 116

Qy    64 TCPLQTGDVI 73
            ||:|||
Db   117 YI-LKNGDRI 125
```

RESULT 9

S73738

phosphoglycerate kinase (EC 2.7.2.3) pgk - Mycoplasma pneumoniae (strain ATCC 29342)
C;Species: Mycoplasma pneumoniae
N;Alternate names: hypothetical protein A05_orf409
A;Variety: ATCC 29342
C;Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C;Accession: S73738
R;Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A;Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A;Reference number: S73327; MUID:97105885; PMID:8948633
A;Accession: S73738
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-409 <HIM>
A;Cross-references: EMBL:AE000040; GB:U00089; NID:g1674091; PIDN:AAB96060.1; PID:g1674091
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C;Genetics:
A;Gene: pgk
A;Genetic code: SGC3
C;Superfamily: phosphoglycerate kinase
C;Keywords: ATP; phosphotransferase

```
Query Match           18.8%;  Score  71;  DB 2;  Length 409;
Best Local Similarity  29.2%;  Pred. No. 2.3;
Matches   19;  Conservative  14;  Mismatches  20;  Indels  12;  Gaps   2;

Qy    19 SFPSNKLVSGDHCRIV----------VDEKSGQVTLEDTSTSGTVINKLKVVKKQTCPLQ 68
            |||||||||||| ||          :||||| :|| :: |:|  :  :||||||||||
Db    41 SLDTIKYLLGHNCKIVLLSHLSRVKSLDDKKGKKSLQPVASA--LQNLLKNTKVHFCPEN 98

Qy    69 TGDVI 73
          ||| :
Db    99 TGDKV 103
```

RESULT 10

A83872

hypothetical protein BH1777 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 15-Jun-2001
C;Accession: A83872
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058112
A;Accession: A83872
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-230 <STO>
A;Cross-references: GB:AP001513; GB:AP001513; GB:BA000004; NID:g10174345; PIDN:BAB05496.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH1777

```
Query Match           18.7%;  Score  70.5;  DB 2;  Length 230;
Best Local Similarity  31.5%;  Pred. No. 1.4;
Matches   23;  Conservative  13;  Mismatches  30;  Indels   7;  Gaps   4;

Qy     3 LRKREWTIGRRRGC---DLSFPSNKLVSGDHCRIVVDEKSGQVTLED-TSTSGTVINKLK 58
```

```
Db    23 VKKESWSIGRLGKSWKPDIAF-DNVFISRKHALLYVEE--GQVFVKDLDSKHGTYVNDQR 79
           :: |::|||             :  ||    ::  |     ||:|| |::||::|| :||

Qy    59 VVKKQTCPLQTGD 71
           :::||||||||
Db    80 LAPHAPERLSHGD 92
```

RESULT 11

T40402

forkhead nuclear signaling protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Jimenez Martinez, J.
submitted to the EMBL Data Library, August 1998
A;Reference number: Z21926
A;Accession: T40402
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-301 <LYN>
A;Cross-references: EMBL:AL031261; PIDN:CAA20309.1; GSPDB:GN00067; SPDB:SPBC3H7.13
A;Experimental source: strain 972h-; cosmid c3H7
C;Genetics:
A;Gene: SPDB:SPBC3H7.13
A;Map position: 2

```
Query Match           18.7%;  Score  70.5;  DB 2;  Length 301;
Best Local Similarity  27.3%;  Pred. No. 1.8;
Matches   21;  Conservative  18;  Mismatches  27;  Indels  11;  Gaps   4;

Qy     8 WTIGRRRGCDLS-FPSN------KLVSGDHCRIVVDEKSGQVTLEDT-STSGTVINKLKVV 60
          : |:: :     :||||      |:: |:  :|  ::::  |:  :|  | ||||||::
Db    27 YKIGRHTNKSTSPSPSNLFFNSKVLSRQHAEELWLDKDTLSVYIRDVKSSNGTFVNETRLS 86

Qy    61 KKQ---TCPLQTGDVI 73
          ||:   ||||||| ||
Db    87 PENKPSAPCKLNSGDIV 103
```

RESULT 12

S76488

hypothetical protein - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S76488
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S76488
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-157 <KAN>
A;Cross-references: EMBL:D90915; GB:AB001339; NID:g1653604; PIDN:BAA18617.1; PID:d101935
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

```
Query Match           18.1%;  Score  68.5;  DB 2;  Length 157;
Best Local Similarity  32.2%;  Pred. No. 1.5;
Matches   28;  Conservative  11;  Mismatches  25;  Indels  23;  Gaps   4;

Qy     1 VLLRKREWTIGRRRGCDLSFPSNKLVSGDHC--------------RIVVDEKSGQVTLED 46
           ||| :   ||||||  :|||| :  : |||              ||::| ||| |||
Db    25 VLLTETFYTIGRSPRADIRIKS-QFVSRIHAVLVRKSSDDVQAAYRIIDGDEDGQ-----78

Qy    47 TSTSGTVINKLKVVKKQTCPLQTGDVI 73
          |: || ::  ||:| :|||||||||
Db    79 SSVNGLMINGKKV--QEHIIQTGDEI 102
```

RESULT 13
AC3089
conserved hypothetical protein Atu4335 [imported] - Agrobacterium tumefaciens (strain C5
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AC3089
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AC3089
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-399 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL45129.1; PID:g17742801; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu4335
A;Map position: linear chromosome

Query Match　　　18.0%;　　Score 68;　DB 2;　Length 399;
Best Local Similarity　　29.3%;　Pred. No. 4.9;
Matches　22;　Conservative　16;　Mismatches　27;　Indels　10;　Gaps　4;

Qy　　7 EWT------IGRRRGCDLSFPSN-KLVSGDHCRIVDEKSGQVTLEDTSTSGTVINKLK 58
　　　　　:|:　　::|| ||　:|::||　||:|:　|:::|　||||||:||||||:
Db　18 QWSFERGRRAIGRSRDCDWQIDDNERRVSKLHCTLSRDGE-GFIIL-DQSANGTLVDGRL 75

Qy　59 VVKKQTCPLQTGDVI 73
　　　　::|:　:|:　||
Db　76 LLEGSSARLRDGSQI 90

RESULT 14
G98197
hypothetical protein AGR_L_1057 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: G98197
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: G98197
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-399 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89105.1; PID:g15158909; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_1057
A;Map position: linear chromosome

Query Match　　　18.0%;　　Score 68;　DB 2;　Length 399;
Best Local Similarity　　29.3%;　Pred. No. 4.9;
Matches　22;　Conservative　16;　Mismatches　27;　Indels　10;　Gaps　4;

Qy　　7 EWT------IGRRRGCDLSFPSN-KLVSGDHCRIVDEKSGQVTLEDTSTSGTVINKLK 58
　　　　　:|:　　::|| ||　:|::||　||:|:　|:::|　||||||:||||||:
Db　18 QWSFERGRRAIGRSRDCDWQIDDNERRVSKLHCTLSRDGE-GFIIL-DQSANGTLVDGRL 75

Qy　59 VVKKQTCPLQTGDVI 73
　　　　::|:　:|:　||
Db　76 LLEGESARLRDGSQI 90

RESULT 15
T00393
Nijmegen breakage syndrome protein NBS1 - human
N;Alternate names: cell cycle regulatory protein p95
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 20-Jun-2000
C;Accession: T00393
R;Matsuura, S.; Tauchi, H.; Nakamura, A.; Kondo, N.; Sakamoto, S.; Endo, S.; Smeets, D.;
Nature Genet. 19, 179-181, 1998
A;Title: Positional cloning of the gene for Nijmegen breakage syndrome.
A;Reference number: Z14144; MUID:98282099; PMID:9620777
A;Accession: T00393
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-754 <MAT>
A;Cross-references: EMBL:AB013139; PIDN:BAA28616.1
C;Genetics:
A;Gene: NBS1
A;Map position: 8q21
A;Introns: 13/1; 57/3; 107/2; 160/3; 195/2; 234/3; 299/2; 332/1; 375/2; 466/2; 615/3; 63&
C;Superfamily: human Nijmegen breakage syndrome protein NBS1
C;Keywords: DNA repair

Query Match　　　18.0%;　　Score 68;　DB 2;　Length 754;
Best Local Similarity　　24.7%;　Pred. No. 9.8;
Matches　20;　Conservative　21;　Mismatches　30;　Indels　10;　Gaps　2;

Qy　　2 LLRKREWTIGRRRGCDLSFPSNKLVSGDHCRIVVD--------EKSGQVTLEDTSTSGT 52
　　　　　:　|　||　||　:|::||　:|::||　:|:||:||||::
Db　18 LLTGVEYVVG-RKNCAILIENDQSISRNHAVLTANFSVTNLSQTDEIPVLTLKDNSKYGT 76

Qy　53 VINKLKVVKKQTCPLQTGDVI 73
　　　　:|:　:|　||||:||||
Db　77 FVNEKMQNGFSRTLKSGDGI 97

Search completed: May　7, 2004, 14:50:07
Job time : 5.33553 secs